DOCUMENT RESUME

ED 075 030 LI 004 195

AUTHOR	Chernyi, A. I., Ed.
TITLE	Problems of Information Science; Collection of Papers.
INSTITUTION	International Federation for Documentation, The Hague (Netherlands). Committee on Research on the Theoretical Basis of Information.
REPORT NO	FID-478
PUB DATE	72
NOTE	241p.; (101 References)
AVAILABLE FROM	International Federation for Documentation, 7 Hofweg, The Hague, Netherlands (no price quoted)
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Abstracting; Communication (Thought Transfer); Computers; Indexing; Information Needs; Information Retrieval; *Information Science; *Information Systems
IDENTIFIERS	Excerpta Medica; VINITI

ABSTRACT

Twelve papers written by fifteen authors from nine countries comprise this collection which embraces many of the crucial problems facing modern information science and scientific information activities, ranging from information needs of scientists and specialists to the specific aspects of computerization of various information processes. The papers are: Studies of Information Needs; Informal Communication in Science; Some Sociological Aspects of Formal Systems for the Communication of Knowledge; The Aging of Scientific Literature; Conflicting Phenomena in IR Systems; Problems of Data Retrieval And Dependent Techniques; A Condensed Heuristic Documentation System; Proposal and Wishes for an Open Structure in the Communication of Information; The Future of Access (Abstracting and Indexing) Services; Integrated Information Systems; The Activities of Excerpta Medica; and, The Use of Computers in Scientific, Technical and Economic Information in Poland. (Author/SJ)

FILMED FROM BEST AVAILABLE COPY

ED 075030

INTERNATIONAL FEDERATION FOR DOCUMENTATION

PROBLEMS
OF INFORMATION
SCIENCE

FID 478

INTERNATIONAL FEDERATION FOR DOCUMENTATION
STUDY COMMITTEE «RESEARCH
ON THEORETICAL BASIS OF INFORMATION»
(FID/RI)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PROBLEMS OF INFORMATION SCIENCE

Collection of Papers

Edited by *A. I. Chernyi*

FID 478

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
FID
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER"

LI 004 195

ALL-UNION INSTITUTE FOR SCIENTIFIC
AND TECHNICAL INFORMATION

MOSCOW 1972

CONTENTS

Preface
P. Atanasiu (Romania)
Studies of Information Needs 20
A. Merta (Czechoslovakia)
Informal Communication in Science 34
D. Foskett (United Kingdom)
Some Sociological Aspects of Formal Systems for the Communication of Knowledge 53
B. C. Brookes (United Kingdom)
The Aging of Scientific Literature 66
J. Wanke (Czechoslovakia)
Conflicting Phenomena in IR Systems 91
R. A. Fairthorne (United Kingdom)
Problems of Data Retrieval and Dependent Techniques 98
R. Leclercq (Belgium) and *C. Rixon* (United Kingdom)
A Condensed Heuristic Documentation System 111
J. E. J. Halkin (Belgium)
Proposal and Wishes for an Open Structure in the Communication of Information 122
D. B. Baker (USA), *P. V. Parkins* (USA) and *J. Poyen* (France)
The Future of Access (Abstracting and Indexing) Services 142
A. I. Chernyi (USSR)
Integrated Information Systems 167
P. J. Vinken (The Netherlands)
The Activities of Excerpta Medica 208
W. Piróg (Poland)
The Use of Computers in Scientific, Technical and Economic Information in Poland 229

ON CERTAIN KEY PROBLEMS OF INFORMATION SCIENCE.

(PREFACE)

The collection of articles «Problems of Information Science» (FID 478) is the second publication of this kind issued by the FID Research Committee «Theoretical Basis of Information». (FID/RI). The first collection had a similar title — «Theoretical Problems of Informatics» (FID 435) and came off the press in 1968.

The present collection consists of 12 papers written by 15 authors from 9 countries. The collection embraces most of the crucial problems facing modern information science and scientific information activities, ranging from information needs of scientists and specialists to the specific aspects of computerization of various information processes.

The first paper «Studies of Information Needs» is by P. Atanasiu (Socialist Republic of Rumania). It reviews the main methods of studying this important problem of modern information science. In recent years, however, there has been a tendency towards a relative decrease in the number of studies devoted to this problem. For instance, in November 1961 «Current Research and Development in Scientific Documentation» reported 39 projects or studies of information needs and uses; in November 1958—1962, in November 1958—1964, in 1961—1966, in 1943—1969, while the overall number of research projects in information science and technology reported in «Current Research and Development in Scientific Documentation» during this period increased from 271 to 785, i. e. almost threefold.

It can be explained by the fact that academic research into information needs, despite its tremendous difficulty and complexity, has evidently failed to justify the hopes that had been placed on it.

On the one hand, these studies have made it possible to identify the types of information needs and the categories of information users which *had been largely known* by the practical experience in publishing scientific and technical literature and by the experience gained in information and library-bibliographic work. On the other hand, these studies have shown that each particular information need is a very complex and many-sided phenomenon. It was discovered that each information need is, in a sense, unique, for it depends on the user's professional background, his psychological disposition, age, the task facing him and many other factors.

It follows from the unique character of each information need that, in order to raise the efficiency of information service rendered to scientists and specialists, one should strive, wherever possible, to provide the information system with a capability of feedback with the user. Such a feedback system opens up broad opportunities for thorough study of information needs of scientists and specialists. It confirms the conclusion that at *the present stage of information science development the main source of cognition of the regularities of communication in science is the information practice.*

It should also be mentioned that studies of information needs and of their dynamics have been gaining importance as an approach to exploring the processes of creative work, i. e. they are actually outstepping the scope of information science.

The subject discussed in the paper by A. Merta (Czechoslovakia) covers informal methods of scientific communication. For the past 5 to 7 years it has been the main subject of his research. Taking advantage of this opportunity, I should like to express in this preface my own opinion on the general approach to the studies of informal and formal channels of scientific communication.

To begin with, one can hardly agree to any — explicit or implicit — contraposition of informal channels of scientific communication to the formal ones or to declarations to the effect that informal channels are more important than formal ones. When speaking of the *system* of scientific communication we should never forget that both informal and formal channels are organic components of one and the same system. In this system they perform different functions and, therefore, possess different properties.

A most important attribute of a system is that it possesses a certain *integrality,* i. e. that the entropy of the totality of the component making up the system is less than that of a simple aggregate of interconnected and interacting objects and phenomena. If the definitive characteristic of a system were merely the presence of interconnection and interaction between objects and phenomena, any set of these could be called a system, since all objects and phenomena are to a greater or lesser extent interconnected and interacting with each other. In that case, it would be senseless to introduce the concept of «system».

In our view, to maintain that informal channels of scientific communication are more important than formal ones would be as unjustified as to believe that a certain body organ or a system of organs (say, the nervous system) is of a greater importance for the human organism than some other organ or system (say, the heart and the blood system, the lungs or the skeleton).

Explicit or covert contrasting of informal channels of scientific communication to formal ones reveals the metaphysical «sin» of many contemporary scientists and engineers who, while paying lip-service to the dialectical principle of system approach to natural, social or mental phenomena, actualy yield to the temptation of simplifying their task by searching only for a *single* solution, which like a magic number or formula, o «philosophers' stone», Archimedean poin of support, etc. could guarantee the attainment of their goal *under any circumstances.*

In this they sometimes refer to V. I. Lenin's words who wrote: «The whole art of politics lies in finding and taking as firm a grip as we can of the link that is least likely to be struck from our hands, the one that is most important at the given moment, the one that most of all guarantees its possessor the possession of the whole chain».*)

As is evident from the quotation, V. I. Lenin spoke of a concrete approach to phenomena, of the finding of the link which is the most important one *under the given conditions*. It needs not necessarily be as important under different conditions. Thus, this injunction by Lenin is of a methodological nature and it does not concern the structure or the essence of things and phenomena.

---

*) V. I. Lenin. Coll. works, Vol. 5. Moscow, Progress Publishers p. 502

We do not possess a sufficient amount of reliable data to question the assertions of the highly esteemed scientists (D. J. de S. Price, W. D. Garvey, B. C. Griffith, H. Menzel and others) who maintain that at present more scientific information is transferred through informal than formal channels, and that such information is more valuable for science. Moreover, we are ready *to accept* such assertions, although they are based rather on intuition than on scientific evidence. That does not, however, grant the conclusion that at present informal communication channels play a more important role in science than formal ones. Such a conclusion might be particularly dangerous when the administration of science and its financing are fully or to a great extent exercised by the State.

Being interested in obtaining a maximum «scientific output» to justify huge sums allocated for science, the State, with the help of financial and other measures, strives to create conditions which, according to its official experts, ensure the best possible facilities for effective research and development.

Inadvertent — and utterly wrong — assertions of an unconditional primacy of informal channels of scientific communication over formal ones may lead (if they have not already led) to a cut-down of allocations for publishing scientific and technical literature and thus may cause serious damage to the development of science.

Whatever the quality of informal channels of scientific communication, their personal nature and spatial limitations preclude their being capable of ensuring the genuinely international character and historic continuity of science. These are two of the many functions of scientific and technical literature which in the social mechanism of science is, at least, as important as informal channels.

Thus informal channels of scientific communication should be studied in their organic unity with formal ones, so to speak, *in vivo*. These studies should deal not only with the specific features, and properties of informal channels but also with the mechanisms of interaction between informal and formal channels. It is these mechanisms that are among the least studied aspects of scientific communication, although they are likely to hide great potentials of boosting the efficiency of scientific information transfer and use.

There is one more remark. A. Merta asserts that the system of scientific communication is a closed one. We cannot agree with this, for we maintain that the scientific communication system, just as all or almost all social systems, is an *open system*. The hierarchical nature of its structure is the means ensuring the preservation of the system's integral nature (closed systems preserve their integrality due to a feedback with the environment). We believe that the hierarchical structure of social systems explains the amazing circumstance that Zipf's rank distributions can be constructed for the elements of many different social systems. Bradford's law of scattering describes one of such distributions relating to scientific and technical literature.

The paper «Some Sociological Aspects of Formal Systems for the Communication of Knowledge», contributed by D. Foskett (UK), is a brief review of the development of formal methods and means of scientific communication in their totality, which came te be more or less fully denoted by the notion of «scientific and technical literature». The paper reflects great scientific erudition of the author, his rich practical experience, profound understanding of the subject, and the ability to express his ideas in a clear-cut and perfect style. I am sure that the readers will share with me the great interest and pleasure that I felt each time when reading D. Foskett's paper in compiling and editing this volume.

One of the ways to overcome the state of crisis that has been taking shape in the sphere of scientific and technical literature because of the rapid growth of the number and size of periodicals is, according to D. Foskett, the so-called «Bernal plan». It will be recalled that the plan was proposed by J. Bernal at the 1948 London Conference on scientific information and consisted in the following: the practice of publishing articles in scientific and technical journals should be completely or partially abandoned, and manuscripts of these articles should be deposited in special information centres. The duty of such centres should be to promptly notify scientists and specialists of all the incoming manuscripts which may be of interest to them and to produce copies of these manuscripts on request.

It should be noted that similar proposals had been made earlier (for instance, in 1933 by Soviet delegates to the International Geological Congress and also by Watson Davis). But it is thanks to J. Bernal that the proposal became known to broad groups of scientists and specialists and, therefore, it was justly called the «Bernal plan».

Now there are possibilities for technical implementation of the plan. Moreover, to an extent it is being brought into effect in the USSR. But the experience in deposition of manuscripts shows that if these are not «sifted» through a preliminary reviewing, this is bound to result in lower quality of scientific and technical literature and its devaluation which will be harmful to the general progress of science. On the other hand, if such manuscripts are preliminarily reviewed, will this not sacrifice the main advantage of depositing documents before publication — the reduction of time necessary to bring scientific information to the user's notice? Will the organization of a system for reviewing deposited manuscripts (naturally if the reviewing is done as thoroughly as that of manuscripts to be published) not lead to the emergence of technical structures more or less equivalent to editorial boards of journals? These and many other questions should be given a more or less clear answer before accepting or rejecting the «Bernal plan».

We believe that the stand taken by some periodical publications is more promising. They carry only abstracts of (all or most) articles while the full texts of these — in the form of separate off-prints or copies on microfiches — are disseminated on readers' request.

D. Foskett justly criticizes the «engineering approach» to problems of scientific communication which virtually disregards social aspects of these problems reducing these to the category of common mass-production problems. The author also warns against the dangers that attach to the fact that in capitalist countries commercial publishers exercise an ever tighter control over publication and dissemination of scientific and technical literature. It leads not only to a rapid rise of prices for scientific and technical literature (during the past 5 years they have gone up by 10—15 per cent per annum) but also to a degradation of its scientific standards. To sum up, there are hardly any reasons to doubt that scientific and technical literature — periodical publications, books, patent specifications, etc. — still is (and will long continue to be) the principal medium of communication in science. The particular material form in which scientific and technical publications will be issued, say, in some 10 or 15 years from now — whether as conventional bound books or on microfiches, or in some other form — is in this case irrelevant. What is essential is that the institution of reviewers be kept intact and operational. We believe the latter to be an important attribute of scientific and technical literature viewed as a social system, for it keeps the scientific quality standard of publications at a certain level achieved by science at the present stage of its development.

Therefore, the identification and study of the properties and regularities of scientific and technical literature are of very great theoretical and practical importance, as this opens up broad prospects for raising the efficiency of its function of scientific information transfer (dissemination).

Another very important property of scientific and technical information is *aging*. Aging of scientific and technical literature in this case means a decrease in intensity (frequency) of the use of publications by scientists and specialists with the increase in the time which elapsed since the date of publication. To measure the obsolescence rate, R. E. Burton and R. E. Kebler proposed to use the so-called «half — life» period (i. e. the time during which a half of the publications of a certain set ceases to be used). This period is obviously different for various sciences and disciplines as well as for various types of publications. The knowledge of the regularities of the obsolescence of scientific and technical literature makes it possible to work out rational acquisition policies of libraries (to be more exact, of systems of libraries), to optimize the file sizes of documentary information retrieval systems (IRS), etc.

The paper by B. Brooks (UK) who has successfully studied various properties of scientific and technical literature during the past few years, is devoted to the problem of aging.

This work excels in clarity and depth due to the fact that B. Brooks has a perfect command of methods of mathematical statistics and uses them skilfully. They invariably evoke great interest among all specialists in information science.

The article by J. Wanke (Czechoslovakia) deals with such parameters of documentary and management IRS as the frequency of the use of the file, the average storage time of a record in the IRS, the average record length and the average value of a record (a document). The author shows that the third-generation computers having a high-capacity storage with random access are thus far more suitable for the implementation of management IRS, which are characterized by a high file call frequency, a comparatively short storage time of records, a shorter record length and a high value of records. According to J. Wanke, it is due to these properties that the management IRS are developing faster than the documentary ones.

However, when examining the reasons of these phenomena one should also bear in mind other important factors. From a semantic point of view the search of documents is by far more difficult than the search of business information, since in the latter case there are no problems of an adequate linguistic expression in an information request of the actual information need; there are no problems of expressing the main subject of the document in its search profile or of establishing the criterion of relevancy between the document profile and the information request. This is likely to explain (at least, to some extent) the fact that users of centralized information systems strive to develop their own IRS, which enable them to take a better account of the specific nature of their information needs and achieve better retrieval results.

Besides, documentary IRS do not generally yield the same direct and tangible economic effect as management IRS, for instance, the computer-based system for booking air tickets. Last but not least, public practice demands a much greater number of management IRS than documentary ones.

R. Fairthorne (UK) has written for this collection the paper «Problems of Data Retrieval and Dependent Techniques». R. Fairthorne considers data retrieval a special case of document retrieval, as the former seeks to find parts of documents (texts), and document retrieval — complete documents. That this viewpoint is well-grounded becomes particularly clear if one recalls the ambiguity of the notion of «document».

It is not difficult to give examples showing that under different conditions one and the same text can be considered either a separate document or a part of a larger document. In case of data retrieval the user gets texts which contain the answer, or, which is more typical, from which the answer to the given question can be logically derived. The finding of this answer (or logically deducing it) is the task of the user himself.

R. Fairthorne points out that the estimation of the truth or falsity of statements, verification of facts and their use are beyond the scope of information science. This should be always minded by those militant information scientists who strive to include into the subject scope of «information science» as a scientific discipline the content analysis and synthesis of information. These operations are impossible without a qualitative evaluation of information, and the latter presupposes the knowledge and use of theories and methods of relevant sciences and disciplines.

Thus, data retrieval is a sort of continuation of document retrieval: at first the documents are retrieved that might contain parts of the text with the data needed and then — these particular parts of the text. There is nothing peculiar about data retrieval that would substantially distinguish it from document retrieval in scientific, methodological or engineering terms. Whether data retrieval (as well as document retrieval) is used in natural sciences or in the Humanities does not add to it any specific character.

In his paper R. Fairthorne examines most important and pressing questions which today preoccupy many information scientists. We are confident that this thorough and pithy article will be appreciated by the readers.

The paper «A Condensed Heuristic Documentation System» by R. Leclercq (Belgium) and C. Rixon (UK) develops the concept of the language and logic of a formalized logical information system designed for uniform reduction of scientific information as well as for making logical inferences from it and for evaluating new information, not explicitly contained in the input texts (documents).

In this connection we should like to note that we believe that the study and development of logical information systems in the strict sense of the term belongs not to information science but rather to the special sciences and scientific disciplines concerned, for any logical information system is founded on a wide use of a system of concepts, logic, theories and methods of a certain particular science or scientific discipline in deriving and estimating new concepts and obtaining new scientific information. Information science as a scientific discipline does not deal with the content-based processing of scientific information.

Likewise, mathematical logic, for instance, does not intrude upon those fields which serve as substantial interpretations and models of formal logical systems (calculi) it studies. It is evident that information workers who are specialists in relevant sciences and scientifc disciplines, carry out a content-based processing of scientific information when they abstract, annotate and index texts (documents), translate them into another language, write surveys, etc.

J. Halkin (Belgium) has contributed the paper «Proposal and Wishes for an Open Structure in the Communication of Information». This extremely interesting paper concerns many important issues of modern information science and scientific information activities. Among these issues there are some on which we should like to express our views.

J. Halkin considers that at present one of the principal ways of raising the efficiency of scientific information service given o scientists and specialists s decentralization of the processing of documentary information and centralization of its transfer (dissemination). In other words, analytical and synthetic processing of scientific documents (i. e. their abstracting, indexing, etc.) should be done by specialists in narrow fields of science and technology. The results of this processing are transferred to relevant information centres which disseminate scientific information by issuing secondary publications, magnetic tapes with indexed document files, etc.

If I understand J. Halkin right, such a demand — being certainly correct in essence — may cause some bewilderment. For it is precisely the way in which all more or less big information centres of the world really work — VINITI, Chemical Abstracts Service and BioScience Information Service of Biological Abstracts (USA), Centre de Documentation du C. N. R. S. (France), Excerpta Medica Foundation (Netherlands), and others.

At present, over 23,000 *part-time* specialists work at VINITI, many of them living not in Moscow but in other towns of the country. It is these narrow specialists that carry out (in a decentralized way!) the content-based processing of scientific information while VINITI is engaged in centralized fulfilment of all those functions which J. Halkin unites under the common title of «information transfer».

Perhaps, J. Halkin means the necessity of preventing wide duplication of analytical and synthetic processing of the same documents for different branches of science and national economy. But such duplication can largely be accounted for by technological reasons and can only be eliminated as a result of wide application in information practice of computer-based integrated information systems (the papers by A. I. Chernyi and P. Vinken are devoted to this subject). Naturally, the use of computers for the processing of scientific information and for the preparation of secondary publications entails a considerable change in the structure of the networks of information services. If this is what J. Halkin means, then we have no point to argue about.

J. Halkin touches upon one more question of principle which cannot be left without comments. J. Halkin shares with some scholars the view that scientific information is a *commodity*, to which we cannot possibly agree.

What are the attributes on the basis of which scientific information is regarded as a commodity? Such attributes are likely to be the following:

1. Scientific information is the product of labour.
2. Scientific information is bought and sold, hence the conclusion that it has value and a price.

Let us compare these attributes with the most important attributes of a commodity.

A commodity is the product of labour which

— meets a certain human need;

— is produced not for one's own consumption but *for sale.*

Scientific information meets human needs in knowledge, i. e. it is use-value. But not every prмoduct of labour which is a use-value is a commodity. «To become a commodity», K. Marx wrote in his «Capital», «a product must be transferred to another whom it will serve as a use-value, by means of an exchange».[*] However, scientific information is generally obtained *not for sale.*

If one sees only the external aspect of phenomena, there arises an illusion that scientific information is also an object of huckstering. But. what is in fact bought or sold is not scientific information proper, but *the right to use certain kinds of scientific information for the purpose of gaining profit.* Moreover, only the right to use *new technological information,* and not any scientific information, can be bought and sold. The supreme form of scientific information—the knowledge of the laws of nature, society and thinking—cannot be an object of property at all: they are not patented in any country of the world and are not bought or sold.

The same is also true of scientific information disseminated through scientific literature, i. e. of the bulk of scientific information (though the carriers of scientific information — books, journals, copies of articles from them, patent specifications, etc.) are bought and sold. However, prices for scientific and technical publications obviously are not determined by the value of scientific information they contain: books by A. Einstein, N. Bohr, J. von Neumann and L. Landau are sold at the same price as compilations by unknown authors, and any specification to a U.S. patent, regardless of the value of the patented invention, costs 50 cents.

Even if any kind of scientific publication could be bought and sold that would not be sufficient to call scientific information a commodity. For instance, works of art are bought and sold but they are certainly not commodities. Naturally we are speaking in this case of the originals of works of art and not of their copies.

---

[*] K. Marx, Capital, Moscow, Foreign Languages Publishing House, p. 41, Vol. 1.

In his book «The Human Use of Human Beings» (1954) N. Wiener stressed that mercantile attitude to information as a commodity, according to a standard American criterion, leads to the misunderstanding of information and its associated concepts. N. Wiener wrote: «What makes a thing a good commodity? Essentially that it can pass from hand to hand with the substantial retention of its value, and that the pieces of this commodity should combine additively in the same way as the money paid for them. The power to conserve itself is a very convenient property for a good commodity to have <...>.

Information, on the other hand, cannot be conserved as easily, for... the amount of information communicated is related to the non-additive quantity known as entropy and differs from it by its algebraic sign and a possible numerical factor. Just as entropy tends to increase spontaneously in a closed system, so information tends to decrease; just as entropy is a measure of disorder, so information is a measure of order. Information and entropy are not conserved, and are equally unsuited to being commodities.»*)

And further on: «The idea that information can be stored in a changing world without an overwhelming depreciation in its value is false.»**)

It is not appropriate here to dwell in greater detail on the question of why scientific information is not a commodity. We have done it elsewhere. We should like to emphasize, however, that declaration of scientific information a commodity inevitably brings into the scientific information activities a commercial spirit which is utterly inacceptable to all thinking people and which is having the same detrimental influence on it as on similar fields of human endeavour — on science, education, art and public health.

The paper «The Future of Access (Abstracting and Indexing) Services» contributed by D. B. Baker (USA), P. V. Parkins (USA) and J. Poyen (France) is of great interest for, at least, the following reasons. The authors forecast the development and improvement of secondary publications which serve and will long continue to serve scientists and specialists as the main guide through the changeable ocean of world scientific and technical literature. Therefore, the elaboration of the forecast of the future development of this important field of scientific and information activities is of tremendous theoretical and practical significance. The paper is particularly interesting as its authors are well-known experts of great international renown in the field of information science and scientific information activities.

---

*) N. Wiener. The Human Use of Human Beings. London, Eyre and Spottiswoode, 1954, p. 116

**) Ibid., p. 120

Since 1958, D. B. Baker has headed the Chemical Abstracts Service (CAS) which during the last 12 years has developed and introduced into practice quite a few genuinely revolutionizing ideas and methods of automated processing of chemical information. Since 1907, CAS has issued the «Chemical Abstracts» — the biggest abstract journal on chemistry and chemical engineering published in English and known to every chemist. In 1971, this journal published 350105 abstracts of articles, patent specifications, scientific and technical progress reports and other scientific documents, which is a 13 per cent increase as compared with 1970. Since its start, the «Chemical Abstracts» has published a total of over 5 million abstracts. Besides, CAS issues a number of other secondary publications. CAS has also developed and put into operation a computer-based system of searching for chemical compounds — the so-called Chemical Registry System in which the names and structures of some 2 million different chemical substances had been recorded by 1972. At present, CAS is one of the most automated world information centres. Obviously the great experience accumulated by CAS during the 65 years of its existence has been reflected in the paper.

P. V. Parkins has been the director of the BioScience Information Service of Biological Abstracts (BIOSIS) since 1965. Beginning from 1928 BIOSIS has issued a leading biological abstract journal in English — «Biological Abstracts». In 1971 this journal published 140,020 abstracts of articles, books and other scientific documents. Besides, in 1971 the «BioResearch Index» carried information on another 100,000 progress reports on research in biology.

During the last 6 years BIOSIS has scored great successes in the development and improvement of their secondary publications and services as well as in the automation of the technological processes used by it.

Thus, P. V. Parkins' ideas of the future of abstracting and indexing services are based on her great practical experience and the results of fundamental research.

The third author of the paper «The Future of Access (Abstracting and Indexing) Services» is J. Poyen who up to 1963 had been engaged in developing automated information retrieval systems for Compagnie des Machines Bull. Since 1963 she has been Secretary-General of the Abstracting Board of the International Council of Scientific Unions (ICSU AB). This international organization was founded in 1952. At present, its members are 13 major abstracting services of the world including VINITI's Abstract Journal (USSR); «Chemical Abstracts», «Biological Abstracts» and «Engineering Index» (USA); «Bulletin Signalétique» (France); «Physics Abstracts», «Electrical and Electronics Engineering Abstracts» and «Computer and Control Abstracts» (UK) and others. In recent years ICSU AB has carried out a number of research projects aimed at developing and improving abstract journals in major branches of science, and has played an important role in undertaking a feasibility study on the establishment of a World Science Information System (UNISIST). Therefore, J. Poyen's thoughts on the ways of further development of abstract journals are worthy of great attention.

Generally, we share the views on the future of abstracting and indexing services set forth in the paper by D. B. Baker, P. V. Parkins and J. Poyen. But we should like, together with these authors, to emphasize once again that the introduction of computers into the practice of scientific information activities entails a radical revision of all traditional concepts, a partial or even complete rejection of the existing stereotypes of technological thinking. One can sometimes observe attempts at using computers to automate the same technological processes that had been practised in the conditions of manual or mechanized labour. This is a capital error. Such attempts can be compared with the behaviour of an airline manager who would so chart the route of planes carrying loads and passengers in a mountain area along the motor roads running through that area that it would follow exactly every meander of these roads.

The real value of the computer is not that it permits to replace manual work by machine operations and to greatly speed up the performance of these. The computer provides prefectly new opportunities that are unthinkable with manual or mechanized labour. As J. Diebold noted, automation should be viewed as a phenomenon causing a change of the basic principles of production. Automation enables the whole production process to be performed as an integrated system, from the input of raw materials up to the packaging of the product. He wrote that automation is «more than anything else a concept or a way of approach in solving problems and that it marks a considerable departure from many accepted practices of management».*) And further on: «Automation is not a particular group of new machines. It is a new concept — the idea of the self-regulating systems — and a new set of principles. Only when management understands this will it gain the full benefit from automation».**)

Automation demands a complete *re-conceptualization*: it is a new way of thinking which permits to avoid the difficulties arising in using the traditional methods of problem solving, and to approach the problem from a new side. This approach should be adopted in designing any kind of system, including an information one.

The paper A. I. Chernyi (USSR), «Integrated Information Systems», defines the concept of an integrated information system, discusses the major factors and premises which led to the emergence of the systems of that type, and gives their generalized functional scheme. An overall description is made of the general approach to the creation of such a system on the basis of VINITI, the current status of its development, and the author's ideas on the possible structure of a nationwide system of scientific and technical information in the USSR to be based on a network of interacting integrated information systems, «information banks» and scientific document reproduction centres.

---

* J. Diebold. Beyond Automation. New York, Mc Graw-Hill Book Co., 1964, p. 54.

** Ibid., p. 55.

The next paper by P. Vinken describes an integrated information system which is already operational — the automated medical information system that was developed at the Excerpta Medica Foundation (Netherlands) in 1968. The author is Director (for Research) of the Excerpta Medica Foundation. His paper is, therefore, perfectly competent on the subject and is extremely interesting, particularly in view of the scarcity of published factual data on this — in a way, unique — computer-based information system.

The book ends up with the paper by W. Piróg, Director of the Central Institute of Scientific, Technical and Economic Information of Poland (Warsaw), entitled «The Use of Computers in Scientific, Technical and Economic Information in Poland». The paper surveys the major accomplishments of Polish specialists concerned with the introduction of computer facilities into scientific and technical information work. It contains much interesting factual information unlikely to be known to the broad audience of information scientists and officers outsido Poland. The paper affords an insight into the basic approaches of Polish specialists to solving this key problem, which makes it especially valuable.

So much for the contents of the collection. In conclusion, we should like to express the hope that this book will contribute to defining more clearly the subject of information science and its main objects of research, to identifying the key problems facing information science and, in this way, to promoting its shaping as a scientific discipline in its own right. This is the main objective of the FID Rsearch Committee «Theoretical Basis of Information»

A. I. Chernyi

P. Atanasiu
(Romania)

STUDIES OF INFORMATION NEEDS

1. Causes that led to the study of information needs

The information system acts as an optimization relay in transmitting knowledge from information sources to users. The major objective of the whole complex of components and activities characterizing this system is to provide the information required at the right time, in an assimilable amount and in a readily accessible form. Given all this, one might wonder why studies of these needs and requirements, underlying any information system, began to develop as late as 1948 [1] and went on concentrating almost exclusively in one country until the 1960's. This was most certainly due to the difficulties encountered in such studies. The numerical growth of users as creators of information, coupled with the proportional increase in the bulk of document calls for such huge efforts in terms of highly-qualified manpower consumption and equipment costs that the question of looking into the needs of information has lately become the focal point of modern information science everywhere. Any mechanized information system requires, before being launched, the acceptance of its prospective users, i. e. the assurance that it will be used and will operate on a profitable basis [2, 3].

Using available data from literature and over own experience in this field, we will try to outline the major factors inherent in any attempt at an information needs analysis in terms of sources, objectives, working methods and techniques as well as prospects.

2. Data-gathering sources

As information creator and user, the specialist, whatever his scale or field of interest, is the main source of data to be taken heed of when evaluating the needs of information, or, to put it more plainly, the requirements vis-à-vis the information system.

The diversity and subjectivity of the users' viewpoints and of the factors acting upon the needs and requirements of information made some authors [4] to question the utility of users' needs investigation in information system design and others to consider the task of defining what information needs amount to as a painstaking or even impossible venture [5, 6]. On this line, there are some supporting the view that information is a professional service which must impose on the user what it deems useful [7] as is the case with the medical sciences.

Still in its early stages, the information science has not yet accumulated a sufficient amount of experience and virified data to be in a position to establish, with the required certainty, what can be of topical relevance in the creative process out of all recorded knowledge. An analysis of the needs of information users could make it possible to thrash out in the foreseeble future some objective and definite criteria to help in assessing beforehand the relevant pattern of information for each separate case.

Deprived as it is from physical attributes, informations are liable to such subjective interpretations that the differentiation between information requirements goes down right to the individual user [8, 9, 10]. It goes without saying that no present-day information system can be altuned to individual requirements and this holds true for the future as well.

Modrn scientific developments by team work and the identical characteristics of larger user communities make it possible for the requirements vis-à-vis the information services to be controlled at a fairly general level [9, 11, 12]. These average needs and requirements must be taken into account in any user studies.

The users' overt reticence and passivity about the information system and most particularly about any breakthrough in this field is common knowledge [12, 13, 14]. The user should be directly involved in the information system operation in order to adapt himself to the technicalities of the user—system interaction. Getting to know the economics of the information system the user will succeed in gauging its advantages and will better understand his duties when using the system.

Users' viewpoints reflect in all cases habits, practices, wishes and needs in proportions hard to determine, the techniques employed in evaluations being often rather uncertain [7].

In spite of all this, the information system must be very sensitive to the users' appreciations whatever their rate of uncertainity if the user is to resort to the information system [3, 6, 8, 9, 16]. User studies must be oriented towards obtaining value judgements and determining some hierarchizations liable to interpretation [17] which should supply the system with the necessary guidelines.

Data obtained from the «user» source must not be taken isolated. They are completed by and correlated with those resulting from the system's technical and material capability as well as from the goals and objectives of the institutions served by the system [18].

3. Goals to be attained in the study of information needs

The specialist in its capacity as creator and user of information and also as main source of data presents several aspects which are often rather difficult to isolate.

Having a marked psychological nature two categories of interests are met with in all user studies: those referring to analysis of the creative process proper (where aspects of the intervention of recorded knowledge are sure to be present) and those regarding the information flow and its implications. The first category constitutes one of the main objects of study of the science of science with corresponding implications for the transfer of information, the second represents the principal object of study in the field of user information needs [19].

Altuned to the problems likely to interest or influence the information system, the study of information needs concentrates on the following: user behaviour in information transfer, channels of information and the ways in which the information flow is put to use [17, 20] and also on the factors heaving upon them.

Researches relevant to the system have to be oriented towards practical problems and designed to gather in addition to descriptive data some indications on the functional relations between user behaviour and information transfer and also some value judgements [17].

One to the practically unlimited amount and diversity of knowledge and to the huge number of users and the changing complexity of their requirements the information system must tackle a wide range of factors in a permanent dynamic interaction. The question of information needs calls for an approach on the complex systems line [21].

Analysis of the principal operations performed by an information system: 1) fields covered; 2) building of holdings; 3) document selection; 4) condensing; 5) indexing; 6) cataloguing; 7) storage; 8) retrieval and 9) display, shows that users' needs bear primarily upon the first three and the last two of the above mentioned operations [21, 22].

Hence the investigation is oriented towards: — establishing personal factors, viz. users' characteristics and their influence on the respective processes;
— determining possibilities and limitations inherent in the needs of information following the intervention of external conditions specific to the social, political and economic system;
— analysing the user-communication channel interaction;
— differentiating between the needs of information according to the latter's functions;
— performance assessment (evaluating performances).

Researches conducted so far generally confirm the afore-mentioned objectives (see table 2) and literature [21, 23, 24] spotlight their utility in information system optimization.

4. Methods and techniques employed in the study of information needs

Bearing in mind that the main data provider is the user himself, the methods employed in such studies, just like those in psychology and sociology are based in the first place on combining introspection with observation

[8]. They are recording habits, practices, preferences and appreciations and of late (after 1960) the effects of information activities. Consequently, they are more of a descriptive nature. Following recent developments investigations of an experimental nature have also come to the fore (see table 3).

First studies covering the 1927—1947 period are centred exclusively on determining the value of publications required by the need of library holdings selection following the information explosion. We may safely say that the study of information needs was initiated by J. D. Bernal in his «Preliminary analysis of pilot-questionnaires on the use of scientific literature» presented at the Scientific Information Conference held in London in 1948 under the sponsorship of the Royal Society [2].

The rather different working techniques — bibliographic reference and library statistics, inquiries, interviews, dairies, case studies, critical incidence studies, system experimentations — are generally based on the forme and questionnaire method [8, 2]. Due to their limitations these techniques are most often combined.

The data gathered mainly refer to the sources, practices and conditions of information. The utilization of the new techniques such as the critical incidence (effects of some information actions) and the system experimentation technique are mostly intended to obtain indications on the uses and results of information actions likily to contribute towards defining and objectively differentiating the needs of information and determining more reliable assessment criteria than the more use frequency of hierarchization.

The intervention of modern electronic means in information enhances the possibilities of data processing and interpretation in this field of research [4] and makes it possible to continously check on the efficiency of the services rendered in terms of users' reaction.

5. Results of the study of information needs

Oriented towards information system optimization, the study of information needs aims at establishing a correlation between what is useful for creation and what the user wants to get on the one hand and the system performance and the factors influencing the mentioned aspects on the other.

Despite the difficulties encountered in such studies and despite the fact that their limited number has not made possible, for the time being, a wider approach of the whole complex of problems*, the efforts made so far have led to results that proved to be of real use to the system. This gives hope and justification for continuing and improving them.

On the line of determining the factors that bear upon information needs researches have taken into account several aspects. It has been generally considered that the user's activity is the main factor inherent in information needs modelling. Hence, the following activities have been listed as involving characteristic requirements: fundamental research; applied research; design; production; education; management and administration [25]. The purpose for which informations are being used, relative to their function, although an outstanding factor by its action on the nature of information has been dealt with starting from the information channels and by determing the functions which are better served by each one of them. This differentiation should be made the other way round starting from the function that must be performed and establishing the optimum sources and channels for its achievement [17]. For all this a differentiation of the functions performed by information has been reached which will make it possible to delve deeper in this field.

Grouped according to the activities they serve the following functions come out: a) professional interests that change with the work stage, i. e. — finding of new ideas (selection of the course of action); — orientation towards problem approach (work planning); — classification of theoretical aspects; finding of concrete data and solutions for the problems raised while working on the project and of informations in related fields relevant to the project (the working stage proper); — synthetizing and fixing the results; b) current awarness as regards achievements and trends in the field of specialized interest; c) orientation towards new disciplines or problems; d) editing of a paper; e) preparation of a lecture; f) elaboration of a conference [19, 26, 27, 28, 29].

---

* The fields dealt with were mainly enginnering, chemistry and physics, the focus being placed on fundamental research and in the last decade on applied and production research, the samples investigated being put to around the one thousand mark [2]

The field of specialized interest has an additional influence on information needs and requirements depending on its nature (whether it is — theoretical or experimental), on the novelty and topicality of the subject matter, namely, on the rate of development, nature and relevance of its relations to other specialities.

As regards the conditions of information, viz. the information environment, the latter acts according to the capability of coverage of the volume of recorded knowledge, the economic possibilities of acquisition und publication, the character and size of the institution, and the system of activity organization (by large or small groupings with wider or restricted participation on the part of specialists from other fields).

Though of a more subjective nature, the personal factors that influence the needs of information have been fairly extensively surveyed. This was also due to the fact that the data provided by the user in this field are far more reliable. The interpretation and correlation of these data with the capability of the information system and with other factors still lies at the bottom of the scale. Among these factors one can list: professional training, qualification, experience, know-how in the subject matter and in related fields, age, knowledge of foreign languages.

The effects of these factors are manifest at both individual and community level. It has been ascertained that starting from fundamental research toward applied research and production, just like starting from theoretical fields to the factual and experimental ones brings about a more marked dispersion of the works, a sensible rise in the role played by the related specialities involved in solving the problems, a lower incidence in the direct use of literature and an upward trend towards recourse to intermediaries (information services, libraries) and to processed materials in a readily accessible and directly usable form [2].

Appart from the above-mentioned factors these findings are also determined by the amount of time available for information that users in the respective categories have on their hands.

As regards the characteristics of information channels an impressive mass of data has been obtained by studies conducted so far. The user knows better this side of the problem since he is constantly confronted with the inflow of information intended to put to use existing knowledge in his own field of interest and with the outflow of disseminating his own scientific creation [30].

Data recorded so far mainly refer to the degree of utilisation of the various categories of information materials and to a lesser extend of the information services. Evaluations and differentiations in point of aims mainly refer to primary literature and direct contacts more familiar to the user than to secondary literature and information services.

Both literature and face-to-face contacts are considered to be of paramount importance in finding new ideas, identifying subject matter information and obtaining concret data and solutions. Literature reigns supreme in the field of discipline-oriented current awareness, in up-to-dateness in a new problem, in elabortaing a paper, conference, etc.

The utilization degree of classical primary sources (periodicals, books) is highly disproportionate as compared to the less usual and never forms (patents, standards, dissertations, technical reports, propectuses) [1, 11], due to external causes (poor accessibility, limited dissemination, etc.) rather than to information contents. Consequently, data on the latter category of documents are still inconclusive and cannot therefore stand interpretation.

Existing studies on primary publications provide fairly reliable data with regard to their importance, viz. value, topicality, lifetime, use frequency [17, 31]. Such data are extremely useful for orientation in profile of holdings and selection of information.

Extensive data have also been gathered as regards the deficiencies and the limited degree of use of secondary publications. The finding that these are first and foremost the working tools of the information service led to sustained efforts intended to bring them closer to users' requirements.

Insignificant use is still being made of the many different services likely to be rendered by an information system. Lack of understanding of such services deprive the data obtained of their partinence. What does the user want to day from an information service goes far beyond the latter's capability. We may safely say that this most important aspect of the use of the information channels is just being tentatively approached.

In point of fact the data obtained in this field are based on the use frequency, opinions, evaluations and hierarchization of the various information channels. Such assessments imply a high degree of subjectivism and uncertainty. New measuring methods should be experimented as some authors recommend on the line of recording information activities and their effects [17].

Researches oriented towards determining the time consumed for information are inconclusive because they take the information process as a whole and fail to differentiate between the operations. Such differentiated appreciations would be of greater use to the information system.

Conclusions

The study of information needs constitutes one of the crucial problems of modern information science called for by the imperative of information transfer optimization. Out of the wide spectrum of aspects of this problem, those which once solved can favourably influence this transfer should receive the greatest amount of attention on the part of the information scientists.

Analysis of data sources, of objectives and working methods employed in such studies spotlight the complexity of the problem, the uncertainty and subjectivity of data, the lack of precision of the methods and the difficulties in devising objective measuring units. In spite of this researches conducted so far have confirmed that in some cases, results with a fairly good concordance can by obtained. Apart from identifying a number of factors that bear upon the information many data have been established as to the users' behaviour to the various information channels and indications have been obtained relative to the users' differentiated assessments with regard to those factors.

Periodic studies of this kind, in unitary working conditions, on representative samples, with a most balanced coverage of the fields of activity, disciplines, categories of users and geographical areas pursuing improvement of working methods by establishing reliable measuring units and the application of modern computation methods in interpeting the data will lead to valuable results for the information system with wide possibilities of correlating the various aspects of the problem.

Table 1

Distribution by time periods and by countries of user studies identified in literature*

Time period	Number of works		U. S. A.		Britain		Other countries	
	Percentage of total		Number of works	Percentage of total period	Number of works	of Percentage of total period	Number of works	Percentage of total period
1927	1	0,7	1	100				
1928—1937	9	6,2	8	88	1	12		
1938—1947	6	4,1	6	100				
1948—1967	50	35	39	78	8	16	3	6
1958—1967	78	54	49	62,8	16	20,6	13	16,6
Total	144	100	103.	71,5	25	17,4	16	11,1

---

*Data are based on a number of 144 user studies conducted in the 1927—1967 period and identified by us in literature

Account will have to be taken of the fact that most of the data obtained generally reflect habits, practices resulting from given conditions. Experimental methods that reveal the effects of a certain type of information system on user behaviour towards the system and its performance lead to far better results in terms of reliability and objectivity despite the difficulties encountered in such studies they should not be abandoned but rather expanded by all available means.

One should constantly bear in mind that the needs of information are in a dynamically upward trend within the

Table 2

Major topics in user studies

Subject matter	Studies conducted between 1927 and 1967 in percentages	Studies still underway in 1968* in percentages
Recourse to information transmitting channels		
Sources of information:		
Evaluation of primary publications-utilization	28,5	5,4
Studies of primary publications	1,4	
Sources of information identification	15,7	9
Evaluation of abstracts journals	2	2
Evaluation of various information channels	6,4	24
Information practices		
Recourse to libraries and informatio. services	5	22
Personal files	1,4	3,6
Conditions of information:		
Time of information	5	
Requests for bibliographic searches	2	2
Exhastiveness, pertinence		
	67,4	68
Performances of information	17,7	10,9
Function of information	2,1	
Ways of information (general)	12	22

---

* After «Current research and development in scientific documentation», no. 15. Washington, National Science Foundation, 1969, viii, 741 pp.

user — information system and channels interaction, closely following the last developments in science and technology. This multiplicity and complexity of relations leads to a growing interest in the question of information needs and is at the same time responsable for the difficulties encountered in establishing these needs.

Without seeking to exhaust this subject with multiple and diverse implications, the present paper is intended to place a somewhat new emphysis on some of its characte-

Table 3

Dynamics of investigation methods and techniques employed in user studies (Figures illustrate, in percentages, the proportion of the works carried out in the respective decade)

Decade	Methods of recording user practices									Methods of recording the effects on investigation		
	Numeration of bibliographic references	Library statistics	Questionnaire-based investigation	Dairies	Intervieus	Questionnaire and interview-based investigation	Investigation and diary	Case studies	Diary and interview	Experimental investigation based on questionnaires and interviews	Critial incidence (Investigation based on questionnaires. and interviews)	Numbr of elaborated works
1927	100											1
1928—1937	89	11										9
1938—1947	100											6
1948—1957	34	8	24	4	22	4		2	2			50
1958—1967	6	9	32	4	13	9	1	4	2	10	9	78

ristics in its endeavour to contribute towards clarifying a number of aspects and prompting additional interest for the study of information needs as an outstanding factor in modern information systems optimization.

REFERENCES

1. S. Aims. Survey on information needs of physicists and chemists. «J. Doc.», 1965, 21, № 2, 83—112.
2. П. Атанасиу, Н. Греку и Е. Тома. Потребители. Прогноз развития научно-технической информации на 1970—1980. Москва, ВИНИТИ, 1969.
3. R. M. Fishenden. Information use studies. Part I, Past results and future needs, «J. Doc.», 1965, 21, № 3, 163—168.
4. G. Jahoda. Information needs of science and technology — background review. Proc. of the 1965 Congress FID. Washington, D. C., Oct. 7—16, 1965, vol. 2, p. 137—142 Washington, Spartan Books. 1966, 264 pp.
5. O'Connor. Relevance disagreements and nuclear request forms. «Am. Doc.», 1967, 18, № 3, 165—177
6. J. O'Connor. Some questions concerning «information needs», «Am. Doc.», 1968, 19, № 2, 200—203.

7. S. Herner and M. Herner. Information needs and uses in science and technology. In: «Annual review of information science and technology», vol. 2. New York, Interscience, 1967, pp. 1—34.
8. H. Borko, Determining user requirements for an information storage and retrieval system: a system approach. In: «Information Systems workshop. The designer responsability and his methodology». Washington, Spartan Books, 1962, 153 pp.
9. L. F. Carter. National document — handing systems for science and technology», New York, Wiley, 1967, 344 pp. Chapter 4: Document users, pp. 102—112.
10. М. Л. Колчинский и С. Л. Круглов, Научно-техническая документация в информационной системе. Сб. «Научно-техн. информ.», 1965, № 4, 9—18.
11. Д. Е. Шехурин, К вопросу о критерии информационных потребностей. Сб. «Научно-техн. информ.», 1968, сер. 1, № 5, 3—7.
12. E. B. Parker. The user's place in an information system «Am. doc.», 1966, 17, № 1, 26—27.
13. A. Resnick and C. B. Hensley, The use of diary and interview techniques in evaluating a system of disseminating technical information. «Am. Doc.», 1963, 14, № 2, 109—116.
14. E. J. Feinler, C. J. Cook, O. Heinz and C. P. Bourne. Attitudes of scientists toward a specialized information center. «Am. Doc.», 1965, 16, № 4, 329—333.
15. R. R. Stanwood. Some observations on user response to an SDI Systems. Progress in information science and technology. Proc. Am. Doc. Inst., 1966 Annual Meeting, Oct. 3—4, 1966. Santa Monica, Cal., Adrianne Press., 494 pp.
16. Metody pruzkomu potreb a informaci informacnich sluzeb. Informacni studie, № 1—6, Praha, UVTEI, 1968.
17. Review of studies in the flow of information among scientists. Bur. Appl. Social Research, Columbia Univ., Jan. 1960, 103 pp., P. B. 156941.
18. R. W. Coover. User needs and their effect on information center administration. Spec. Lbr. 1969, 60, № 7, 446—456.
19. P. Atanasiu and L. Bogdan, Unele consideratii asupra chestionarelor folosite in studiile de deneficiari. «St. cerc. doc.», 1970, 12, № 3, 237—248.
20. C. P. Dourne, G. D. Peterson, B. Lefkowitz and D. Ford, Requirements, criteria and measures of performance of information storage and retrieval systems. Final Report. Dec., 1961, Stanford research inst., Menlo Park., Cal., 72 pp.
21. E. A. Wall, Retionale for attacking information problems. «Am. Doc.», 1967, 19, № 2, 97—103.
22. В. Моравцова, Методы исследования потребностей в информации в информационных органах. Подготовка потребителей информации и изучение нужд потребителей. Междунар. симпозиум, 11—15 ноября 1969, Бухарест, Ц. И. Т. Д., Бухарест 1969, стр. 17—23.
23. W. J. Paisley, Information needs and uses. In: «Annual Review of information science and technology», vol. 3. Chicago, Encyclopaedia Britanica, Inc., 1969, pp. 1—30.
24. Д. Е. Шехурин. Факторы влияющие на научно-информационную активность ученых и инженеров. Сб. «Научно-техн. информ.» сер. 1, № 10, 9—14, 1969 г.

25. С. А. Дубинская, Исследование потребностей в информации специалистов, работающих в области химии. Сб. «Научно-техн. информ.», 1967, сер. 2, № 4, 3--6.
26. A. Merta. Zprav o vysledcech experimentalniho vyskumu informacnich pozadavka a potreb na pracovistech Ceskoslovenske akademie ved. Praha, 1968, 27 pp.
27. А. Высоцкий, Потребности специалистов в научной информации (общая характеристика состояния исследований). Сб. «Научно-техн. информ.», 1968, сер. 1, № 3, 3—7.
28. Э. С. Берштейн, Об информационных потребностях и качественном преобразовании информации. Сб. «Научно-техн. информ.», 1967, сер. 2, № 6, 8—11.
29. A. Mauperon — Adaptation to user's needs, «Aslib Proc.» 1967, 19, № 7, 232—240.
30. A. E. Bayer and J. Folger. Some correlates of a citation measure of productivity. «Science Sociology and Education», 1965, 39, № 4, 381—390.
31. F. Kertesz. The information center concept. Oak Ridge National Laboratory, Union Carbide Corp., Nuclear Div. for the U. S. Atomic Energy Commission. ORNL — TM 2281 copy 99, July, 1968, 21 pp.

A. Merta

(Czechoslovakia)

INFORMAL COMMUNICATION IN SCIENCE

Introduction

In contrast to technology, which has its outlets opening into the sphere of practical activities, scientific research flows into the sphere of information. This makes the generation of scientific information sources and the modalities of information transfer a key issue of the present-day science.

Information exchange is natural and necessary to the system of science as, indeed, to any social activity of mankind. The organized information flow in science, however, differs greatly from its unorganized, stochastic or spontaneous couterpart. We shall deal here only with «sientific information», i. e. information that belongs to and circulates in the system of science as a purposive communication field oriented toward creative work and its management. We thus disregard the well-known fact that the same «science information» can as well figure in other communication fields, such as journalism, culture and aesthetics, administration, politics, etc.

We proceed from the assumption that a scientist functions alternately as the generator and as user of scientific information, the interactive relations, being very often reinforced by this constant interchange of functions, in particular, during personal contacts at scientific meetings, conferences, symposia, personal correspondence, etc. The social status of a scientist is also of great importance in informal communication. For instance, even a gifted beginner has a much smaller chance to get co-opted to the so-called «invisible college» of authorities in his field than an academician of nigh renown who has a long list of publications in the world's leading journals and single volumes to his credit.

The Mechanism and Function of Scientific Information Circulation

Creative scientific work is normally divided into three fields of endeavour:

(1) collection of facts and data (by experimentation or observation);

(2) generation of new knowledge (by transformation of the data obtained or by abstract deductive techniques);

(3) information work (gathering, evaluation and application of the information obtained by others and communication of one's own findings to the community).

It is the third component that creates premises for the social usability of the results of science. This implies the moral and institutional responsibility of the scientist for taking a direct or at least indirect part in creating the best possible conditions for the circulation of scientific information within the framework of the organization in which he is carrying on his research.

In an earlier paper on the social aspects of scientific and technical information communication (1, pp. 54—63), the author offered a detailed description of the process of generation of new knowledge as conditioned by the constant and occasional comparison of one's own creative work with the general state of human knowledge. Prerequisite to this process is the recording and keeping of information in the appropriate artificial memory. This memory may be either individual or group-based, while its creation may take place through a formal or an informal action. Historically the evolution was from informal to institutional information systems, that came to play in mankind's social mechanism the roles of either measuring or control units, depending on whether they serve creative, managerial, productive or distributive functions.

In the process of generation and transfer of scientific information, the scientist fulfils three main functions, that form three domains of information activities:

(1) The information generator domain.

Here, primary and secondary sources (reports and the corresponding papers, monographs, patent applications, etc.) are generated. These sources are written in the peculiar object-languages of their authors that materially affect the objective aspect of the communications. Information transfer involves a certain conflict between the conceptual universes and the object languages of the particular generators and users of information. It may be necessary, therefore, to build a common «retrieval language» of the formal or informal information system in question for the two communicants and, possibly for their intermediary information officer as well. This language is used in writing appropriately indexed abstracts which form the subject condensates of the relevant primary documents. Such indexed abstracts are stored in the memory of a formal or informal information system. From the memory it is possible to retrieve information relevant to requests by any elements of the abstracts. The information systems as a rule do the hard job of surmounting the language and semantic barriers between the generators and users of information.

(2) The information user (receptor) domain

Depending on his need, the user puts questions to a formal information system or tries to get the information he wants through informal channels. The pertinence of the answers he receives is a function of the degree of mutual understanding attainable through the contact of the differing conceptual universes of the two partners in communication.

(3) The information officer (communication intermediary) domain

In the case of indirect information transfer, the differeing domains of the two communication partners are joined by that of the information officer. As a result, the difficulties of the communication process become still more aggravated and the pertinence of the information dwindles. These difficulties refer mainly to institutionalized information systems.

Formal and Informal Information Flows in Science

Following Garvey and Griffith [2] and Menzel [3], the author has shown in an earlier paper [1] that:

(1) the informal (unplanned) communications in science predominates over the planned (instituional) modes;

(2) Limited access information, both oral and written (graphical) is growing in proportion;

(3) the time intervals between the generation of information and its communication to the user are already too long and continue to grow. The public (formal) information network supplies to its users data 3 to 5 years old. And this at a time when the role of the spatial distance between the communicants has palpably decreased due to the present-day communications facilities;

(4) a significant regularity can be observed in the information flow as regards the transmission channels used by the various categories of persons in certain predictable positions and with predicable priorities. The variables that determine the types of channel employed are many and depend on the scientists' attitudes, and the requirements of the research institutions of funding agencies. In case the existing communication modes are unsatisfactory, the information generator or user may create a new channel. Unfortunately, a majority of scientists nowadays strive to publish their papers in the official scientific journals irrespective of whether it is efficient and economical. This desire is perfectly natural, in view of the considerations of personal prestige and the need for a safeguard against possible priority conflicts.

Every information system is a closed social system with more or less strong and permanent social norms. The relations between its elements are dynamic, that is a change of some of the elements normally leads to some functional change of the other elements;

(5) there are three dimensions, or aspects, in the dynamic interrelations of the information system elements (channels):

(a) **Social dimension.** The scientists themselves create the channels to meet their needs, if the existing channels fall short of it. These new elements affect the existing ones through a change in the scientist's behaviour during information searching and transfer. This behavioural change, in turn, modifies the functions of the rest of the transfer media, and this has a bearing on the other main system parameters, for instance, by changing the norms. A case in point is the growing exchange of preprints and letters between the scientists in a rapidly developing field that cannot afford the delays of official publishing.

Such an intensified preprint exchange between individual scientists sometimes expands to a scale when it becomes necessary to set up a special formal, planned mechanism to cope with it. A growth of this mechanism can bring it to supplant the scientific journal. Simultaneously, formal rules to control the contents of material being exchanged are appearing, more and more restrictions are placed on the participation in the exchange group and eventually we are faced with the appearance of conditions — in terms of authorship, readership, editorship and economics — for the birth of a new scientific journal. This mushrooming of new scientific and technical journal titles is observed also in my country, being a worldwide tendency.

(b) **Economiic aspects.** Since information produced by the system of science grows in bulk faster than GNP, information systems have difficulties to face. Funds begin to be pumped out of the old systems into the newer ones. A case in point is SDI services for highly specialized subject fields that are now pressing the discipline-oriented abstract journals into an unfavourable position. In addition to these direct relationships between the information flow and the money flow in the scientific communication system, certain indirect relations between the latter flow and the various system functions are operating. For example, submitting an abstract of paper to be presented at a future conference is a means to ensure a higher travelling allowance for participating in the conference. A failure to take cognizance of such indirect relations is bound to hamper or preclude a correct assessment of the roles of information system elements.

(c) **Formal aspect.** Both formal and informal communication channels are of importance to science. The two types of channel are marked by clearcut differences that, by their opposition, create the valency of the system. Such valencies lead to the shaping, in the framework of a discipline, of an integrated information system, where the informal subsystem is associated with each of the formal components producing a specific effect in each particular case.

One can characterize the formal and informal channels in the following manner:

1. The formal channels are generally accessible and possess a broad scope of potential users; the informal elements have, as a regular rule, a more limited scope of users.

2. Information disseminated through the formal channels can be stored and retrospectively retrieved; in the informal channels information is often stored but temporarily so that it is difficult to retrieve it later on.

3. The formal channels transmit a relatively more «outdated» information than the up-to-date information transferred through the informal channels.

4. Information transmitted through the majority of formal channels is usually in the document form; that of the informal channels is documented but rarely and not regularly.

5. The formal channels are influenced mostly by the users, whereas the informal information flow is influenced by the initiator of the communication (the disseminator).

6. Every information system is more or less redundant, the results of a research often being transmitted through more than one channel simultaneously, differing by the emphasis laid on this or that particular type of information. Redundancy is less with the informal channels, which transmit the same information in varying interpretations.

7. Interaction is lower with the formal channels than with the informal channels.

Informal information exchange between scientists usually occurs in the form of interpersonal communication. In comparison with the group and mass communications through journals, books, and published reports, the interpersonal communications possess certain advantages, notably:

1. **High speed.** A personal communication or preprint are ahead of the official publication by a few months or even years. It takes three years to publish conference proceedings in Czechoslovakia, whereas in some of the Western countries this is done but in a few weeks.

2. **Selective information dissemination.** This is particularly valuable in basic research, where the specialist scien-

tist needs information covering a wide range of the allied fields. This property is lacked by all formal bibliographical or information tools, including indexes of all kinds. SDI systems for a limited scope of patrons may be regarded as a kind of substitute, especially if they employ a computer for storage and retrieval and have powerful peripheral devices;

3. **Selection, evaluation and synthesis of the information transmitted.** In contrast to a response that the reference information collection or some other formal secondary source or service can give to a scientist, the response to a question posed to a colleague has the following advantages:

(a) the colleague will usually give information and not just refer to the relevant documents;

(b) he will suppress in his answer the documents he considers nonrelevant (there remains, however, the risk of losing important relevant information, the answer largely depending on the individual biases of the person addressed);

(c) the information being supplied has been evaluated and sometimes synthesized by the scientist, the user thus turning to his account the knowledge and qualifications possessed by the answerer;

(d) in case the questioner is a practicioner, he can expect the scientist to communicate not only information but also his opinion concerning its applicability in practice;

(e) the possibility of getting «know-how» information on methods, experimental techniques, instruments, apparatus, that is usually ommited in scientific papers as insignificant and requiring much space to describe; and

(f) direct feedback in exchange of knowledge, that essentially accelerates research and makes it more efficient and saturated. It is a further advantage of the direct interpersonal communications that the interaction helps specify the requests and better to formulate the answers.

Special Role of Casual Information in Science

Occasionally acquired information continues to play a necessary — and thus far not formally described — part in science. Such unplanned communication can occur both in interpersonal and group or mass intercourses. Leaving aside the occasional prompts of the physical world (Newton's apple), most scientists will confirm that they happen to come across exceptionally motivating information while browsing literature or attending a lecture in a field far-off from their own. Menzel (3) notes the phenomenon, as yet unexplored, that the motivational value of information is disclosed only in a repeated receipt of the same information after a lapse of time, in a different creative situation when the receptor has acquired the necessary qualification and experience for revealing the descriptive (methodological) or factual value of that information. This repetition cannot be planned either by the user or the information system that serves him. A classical example is the forgotten inventions that were duly appreciated only years after their first announcement. A possible explanation may be that certain innovative ideas appear much ahead of the feasibility of their social implementation as granted by the intellectual, moral or technological level of mankind at the time they first apear. This shows that occasional impulses, redundancies and other imperfections of the communications systems may happen to be beneficial to society.

1. Interaction of formal and informal communications channels

The two forms of communication have been gradually overlapping of late, especially due to the mechanization and automation of formal communications. This is manifested in an essential reduction of the publication time, organized exchange of preprints among the members of exchange groups, computer-based SDI systems, setting up of special departments for analysis and synthesis of specialized discipline-oriented information (for example, the analysis group at the Central Office of Scientific, Technical and Economic Information), publishing of aids, textbooks and surveys, programmed texts, etc. All this is washing out the sharp borderlines between the two forms of communication. Premises to this development are elaboration of a clearcut conception of the discipline concerned and compilation of good special dictionaries. Coupled with the technical facilities for remote gathering and transmission of knowledge, this opens up unprecedented prospects for building powerful information systems of verbal and non-verbal communication.

The seemingly utopian dream of instantaneous feedback between the communicants has been given the hope of coming true in the information systems affording «user—computer» dialogue. The present-day programming experience attests to the feasibility of machine control and addition of standardized information elements to search prescriptions. Naturally, this mechanized modification of the subject specification of a request at every retrieval phase is conditional on algorithmization and programming of interactive relations between the requestor and the information system, which has been so far a very complicated problem both for the information scientist and the programmer. Among the few papers dealing with this subject one can cite a study by Don Swanson [6].

The efficiency of formal and informal communication channels depends on three factors:

(1) **the intellectual content of the communication,** characterized by the measure of ambiguity of the conceptual vocabulary of the field in question and the symbols and codes used for communication;

(2) the **social structure of the communication partners.** Of major importance in this respect is the size of the group, the number of sources and the degree of commonness of the interests of the communication partners.

(3) **the ideologies of the user group.** Basic differences will be observed in communication between fundamental and «applied» scientists, production engineers, tradesmen, etc. As shown by a field study conducted by the author in the Czechoslovak Academy of Sciences, significant differences are observed also between the social, natural and technical sciences.

In basic research the number of communicants is much smaller than elsewhere, their scope of interests is narrower and better definable, which makes for a certain similarity of the information needs. On the contrary, applied research and production users are marked by a greater diversity of vocations and greater differences of the intellectual standards, which expands the variation range of and deepens information requests. While basic researchers give preference to informal communication and regard the evaluation and synthesis of the contents of relevant documents as inseparable from their creative approach, engineers and other practitioners opt for centralized information services and special libraries. In this latter group the **language barrier** is of a greater weight (see below for more detail).

The **development rates of the discipline concerned** exert a definite impact on the information process. Rapid personal and physical changes in the structure of the discipline usually lead to serious communication handicaps. This fact is well known in Czechoslovakia where science, engineering and the national economy in general are engaged in a continual reorganization. The transition to the federal state structure and to the new economic management methods is bound to entail profound changes in the techniques and organization of the system of scientific and technical information in the country.

For all the advantages of mechanization and automation in the field of formal communication of special information, all kinds of informal exchange will continue to play an important role, particularly, in the sphere of basic research. The effective organization of personal contacts between the eminent and rank and file scientists and equal opportunities of access to information for all the users irrespective of their social status will surely much raise the economic efficiency and efficacy of communication at all levels and stages. Formal communications lack the adequate means to assess the relevance of external information. Informal communication is «open», in that it permits to carry on a dialogue with the communicant, selecting the necessary information, including even such as the user would be incapable of asking for in advance as it is only in the course of the dialogue that he learns of its existence. These specific needs are changing in the process of the problem solution, depending on the nature of this problem, working methods and techniques he applies, the approach to the description of his own results, phase of research, etc. The formal channels are far from always capable of real-time adaptation to the specific and direct needs of each particular user.

The informal channels give the scientist direct support or criticism on the vague points that occur in his work. This feedback in the dialogue with the other members of the learned community in one's field contributes greatly to the efficiency of the research.

It is highly expedient sometimes to combine the two types of channel, e. g. to find a source through a formal channel and then to establish direct informal contacts with the author.

Exchange of preprints has come to be most used informal channel of late. To say nothing of the difficulty of reproduction and circulation of preprints the author has to cope with, this channel involves certain other shortcomings, the major one being that junior research staff, the smaller serearch establishments and scientists in the smaller and less developed countries, who need preprints most of all, have the least chance to obtain them (if at all). Therefore, Garvey [2] suggests as a solution to the problem of lengthy time lags between completion of a research and its announcement in a journal (varying between 1 and 2 years) that lists of manuscripts submitted for press be published with the names and addresses of the authors. In this way informal exchange of preprints would be streamlined to foreshadow the actual publication of the results in the journal. In Czechoslovakia preprints and reprints constitute a major form of informal communication in research institutes. Unfortunately, the scientific journals in the country have so far failed to implement this happy idea of Garrey's of publishing lists of titles and authors of papers submitted.

Publishing of conference pre-papers is among very efficacious semi-formal methods. A special bulletin of brief communications is a novelty in the field. This delivers the official journal for the more important and lengthy publications.

A combination of regular conferences of specialists with the timely issuance of pre-papers is the most effective method of semi-formal communication, thanks to which the participants coming to the conference have a more or less exact notion of the relevant studies so that they can tie up personal contacts with the authors, if they wish. The speakers can thus concentrate on a discussion of their results. The convenors of major international conferences always take pains to publish in due time the materials for the future conference.

Money permitting, after the conference collections of the major papers are published together with the discussions. A tangible percentage of these papers later on appear through the next year or years in scientific journals or monographs at home and abroad. These publications are of particular value to those who did not attend the conference and, certainly, to the information agencies and special libraries that add them to their files.

We can say in conclusion that, as shown by studies carried out in Czechoslovakia and abroad, it is expedient to attack the present information crisis by combining, in accordance with the specific needs of a discipline or a region, the conventional formal channels with the informal ways of communication. Interaction of the two forms is bound to promote the permanent improvement of both. It would be necessary, however, to abandon the intuitive method and to approach this problem with the tools of rational empiric and theoretical analysis. All users of scientific information are eagerly looking forward to its solution.

Influence of the Language Barrier on Formal and Informal Communications in Science

Prerequisite to effective communication is that the symbolic system employed be understandable to all the partners. As regards oral communications, this implies not only the understandability of the unit signs, such as the phonemes and letters of the common alphabet, but that of words, phrases and larger text segments. Otherwise, in case each author uses a different national language to report his information, we are faced with the obstacle of the **language barrier.** In the context of the international nature of scientific information turnover, the language barrier is a big handicap to the effective communication in science and technology.

The problem of language barrier should be explored under three angles:

(a) inter-language barriers (symbolic and linguistic difference);

(b) intra-language barriers (terminological differences of the particular subject fields); and (c) semantic barriers (growth of the degree of abstraction in and conceptualization of the individual fields of science).

In this context we speak of the «big» and «small» languages and of a degree of information pressure that is manifested in the striving to compel the smaller nations to communicate in the language of a larger nation or country. Different to the culturalaesthetic or the journalistic domains of social activities, this politico—cultural pressure is even more strongly felt in the field of scientific and technical information communication. The scientists of the smaller nations and countries that make discoveries of a worldwide purport are practically forced, if they want international recognition for themselves and their nation, to publish and communicate in some language of world importance and to use the international communications channels for this purpose. In other words, they have to publish their results in English, Russian, French, German or international journals. Information flowing through these channels has a greater chance of getting cited by the world's secondary publications, i. e. abstract journals and indexes, thus entering the worldwide turnover.

There are a number of ways to surmount the language barrier in science:

1. **Accepting one of the major languages of the world as the universal international language of science.** The choise of such a language raises complications of psychological, political and prestige nature on the part of the other big and small nations, even though the practice is that without the knowledge — at least passive — of English, Russian, French and German a scientist, particularly in a small country, cannot attain scientific recognition. The majority of special terminologies make use of Latin roots, which to some extent simplifiers the communications between scientists of different nationalities. A case in point is medicine. These circumstances may be the chief reason of the obvious preference that has been given by scientists to the English language in the past two decades, apart from the dwindlings of the international prestige of France and Germany as world powers.

2. **Creation of an artificial international language** of science, a sort of Esperanto for scientific communication

on a world scale. The experience with Esperanto and other artificial languages shows this to be a difficult path as it requires that every scientist devote a lot of his time to learning a new foreign language. Besides, it is difficult in itself to build a universal artificial language for the science as a whole.

3. **Gradual development of formalized languages** for the individual fields of science. Examples are chemistry, physics and other disciplines possessing an articulated systematic ogranization of concepts.

4. **Creation of international retrieval languages** using generally accepted systems of symbols for the individual fields of science and technology. This path is made practical in case there are information agencies functioning as the intermediary of scientific communication. As regards the immediate intercourse of the communicants, a commonly known national language will have to play an important role for a long time to come.

5. **Translation of documents from one natural language to another on user's request.** This is the satisfactory least solution as it considerably retards the information turnover and is fraught with the danger of unvoluntary or intended distortion of the knowledge translated. Unfortunately, this way is indispensable in case the user does not read the foreign language in question. To reduce the disadvantages, some major countries conclude agreements for reciprocal translation of whole series of their scientific journals. For example, about 80 periodical titles in the USSR and the USA the being translated cover-to-cover according to the cultural agreement between the two nations. For economic reasons this solution is unfeasible for the smaller countries.

Linguistics, including the mathematical linguistics, is trying to overcome the language barrier via **machine translation.** The problems involved in the algorithmic description of the translation process, particularly as concerns semantic and syntactical analysis of the source and target languages, will for a long time ahead doom machine translation to the status of laboratory and semi-operating experiments, even though the total number of research reports and various publications on the subject amounts to several thousand.

«Invisible Colleges» and their Role in Science

It has been mentioned that the reciprocal formal and informal communications between scientists are a major stimulus promoting research activities. This communications can take place in the interpersonal plane, according to Schramm's scheme [4], or in the group-based plane, or else, on a limited scale, with the aid of the mass media. In the former case the communication process comprises the following components: the speaker's idea — the semantic content of the message for the speaker — the semantic content of the message — the semantic content for the receptor — and the effect produced on the receptor. In the case of group communication one has to take into account the motivation pattern, the internal horizontal and vertical breakdowns as determined by the roles, norms, statutes and hierarchical relations in the group and its universe. The differences in the statutal hierarchy correspond to the different roles in communication. The senior and managerial staff usually put more initiative in the communication process than do the junior staff.

The above factors lead to the emergence in science of the so-called «invisible colleges». The leading and renowned scientists form, on a voluntary basis, national and international groups without any official organization but with the main objective of information exchange through informal channels, in particular through personal contacts and correspondence, participation in international congresses, conferences, symposia, probation service, exchange of preprints (copies of manuscripts accepted for print), and reprints (offprints of published papers), and so on. These interpersonal or, sometimes, mediated contacts between the members of the invisible colleges elevate these informal groups existing in almost every field of science to the level of mighty informational and brain trusts that constitute a most efficient informal means of communication. The first to study the behaviour of such specific social groups were Price and Beaver [5]. The object of their study was the so-called «information exchange group». There exist in the United States such groups of scientists of different specialties that have undertaken as members of an invisible college to send in to the group's secretariat a copy of each paper ready for press. The secretariat multiplies these manuscripts and disseminates them as preprints among all the members of the group. This simple system has practically delivered the larger groups, uniting up to 500 leading scientists, of the necessity to publish an official journal, for the members of the group received all information on the major studies in their subject field several months ahead of official publication. Naturally, other forms and channels of communication were also formed between the members of the group, as described above, and even direct scientific and technical cooperation. Price and Beaver must be given credit for their quantitative analysis of the social behaviour of the exchange group in the measurable indicators of the publishing activities of the individual group members; besides they examined the related growth trends in joint authorship. Judging by the number of publications, that Price assumes to be a measure of a scientist's productivity all members of the group were ranked in the diminishing order of their publication amounts. Besides, correlation between the productivity and participation in the group was measured, and this dependence was established to be direct and symmetric. The previously known tendency towards the growth of joint authorship was confirmed. A valuable contribution of the research was the extension of the analysis to embrace the relations of the individual scientists to the different groups and the interrelations between the groups. It would be of great use for theoretical research in informatics in Czechoslovakia, if these methods were applied there.

Price, like many other information scientists, resorts to citation analysis to study group communications. Don Swanson [6] makes use of citation analysis even to identify the membership of invisible colleges.

We should like to conclude this brief outline of the scope of subjects connected with the information exchange through invisible colleges, this most efficient channel, by mentioning a major shortcoming of this form. Regrettably, to become a member of such a group is possible only for eminent scientists possessing a social status that enables them to keep sufficiently close contacts with the other members of the group, to take part in conferences, to travel, etc. The choice of the members for an invisible college is quite conservative. A new member is usually required to have high scientific qualifications and renown, confirmed by a multitude of publications in world press, and, in case of a smaller nation, the command of the main languages, as no interpreters are tolerated to mediate the personal communications between the members of the invisible colleges. This problem of the language barrier is of prime importance for the scientists of smaller nations. They need the knowledge of world languages not only for personal contacts, but also for publishing their results in the world national and international journals since, it is only the separate offprints of special world press that grant the entrance to the invisible college. Another opportunity for the scientists of the smaller but developed nations to enter the world scene is given by the current-awareness periodicals of the «Current Contents» type, that give translations of the contents pages of the national scientific journals intc a «world» language accompanied by the authors' addresses — which provides an opportunity for establishing personal contacts.

All junior scientists naturally aspire as soon as possible to become members of an invisible college of their field. The knowledge of world languages, over and above the specialty, is a basic premise to a scientific career in a small country. The capability of overcoming the language barrier on one's own is typical for the field of basic research. On the contrary, in applied research and in the production sphere this capability is lower. This faces the information officer with the major task of bridging the language gap between the foreign sources and the user. Hence the usually higher social status of the information officer in industry than that of the information officer in basic research.

The mass communication media, including the graphic ones, are typical of the formal (planned) communication, and we shall not dwell on them here.

The following table compares the three modes of communication.

Comparison of Characteristics of Interpersonal and Group Communications and Mass Communications

Interpersonal communication	Group communication	Mass communication
Author and ceptor are in- viduals	Author and receptor are either individuals or groups	The source of communication is an individual or an organization, the receptor is the community. Stochastic nature of the communication
Contacts between the communicants ith the medium of structures (se- iantic systems, social roles, norms, ocial status)		Contacts between social structures with the mediation of individuals
Author and receptor are in direct ontact		Author and receptor usually separated in space and/or time
Author addre- ses a particu- ar person	Author addresses a particular person or members of a group	Author addresses a set of receptors (the set may be empty)
Communication is a reciprocal process with alternating roles		No direct (immediate) exchange of roles takes place. Indirect exchange of roles involves a time shift
Physical presence of receptor in he act of communication is necessa- y		Physical presence of receptor not needed
Receptor has no free choice of the ommunication subject in its first hase		Receptor has opportunity of choice of the communication subject (negation of communication)
It is possible to comprehend the ntire communication process		It is impossible to comprehend the entire communication process

Conclusion

The objective of this paper was to emphasize the fac that the informal channels of information exchange i science cannot be replaced even by a best organized for mal communication media in the overall complex systen of the communications model. Informal — specificall interpersonal — communication contacts between scien tists will continue, even at a stage of high automation o scientific information processes.

The need for finding an optimal interaction betwee formal and informal contacts should not be altogethe ommitted from any scientific information project, since i is only optimization of this interaction that can brin about a high efficiency of such a system.

REFERENCES

1. A. Merta, Společenské aspekty komunikace odborných informac Praha, ÚVTEI, 191 pp. (In Czech.)
2. W. D. Garvey and B. C. Griffith. Reports of the America Psychological Association: Project on Scientific Information Ex change in Psychology, Vol. I, II. USA, APA, 1965, 474 pp.
3. H. Menzel, Unplanned communication. In: «Communicatio in Science». London, Churchill Ltd, 169, pp. 57—71.
4. W. Schramm, How communication works. In: «The Process an Effects of Communication». Urbana, 1957.
5. D. J. S. Price and D. B. Beaver, Collaboration in an invisib le college. «Amer. Psychologist», 1966, 21, 11, 1011—1018.
6. D. R. Swanson, Scientific journals and information services o the future. «Amer. Psychologist», 1966, 21, 11, 1005—1010.

D. J. Foskett

(United Kingdom)

SOME SOCIOLOGICAL ASPEKTS OF FORMAL SYSTEMS FOR THE COMMUNICATION OF KNOWLEDGE

The origins of formal systems for the transfer or communication of knowledge are lost in the mists of time, but they were certainly in the hands of rulers, usually kings or priests. Discoveries such as those of the Sumerians and Egyptians, connected with seasonal ebbs and floods of the great rivers, so important for agriculture, were handed on by oral tradition, and, if recorded, were so in great secrecy. The growth of market economies, and the development of international trade, led to an increase in the need for records, of commercial transactions; but the notion that specialist knowledge was the prerogative of the elite lasted long into the history of civilisation, after the invention of the papyrus roll or even of the codex form of book. In his life of Alexander, Plutarch records a letter that he wrote to his tutor while he was in Asia: «Alexander to Aristotle. Greeting. You have not done well to publish your books of oral doctrine; for what is there now that we excel others in, if these things which we have been particularly instructed in be laid open to all? For my part, I assure you, I had rather excel others in the knowledge of what is excellent, rather than in the extent of any power and dominion. Farewell.» The Pre-Socratic philosophers, who made so many important discoveries, and Plato himself, whose influence on the history of ideas can hardly be exaggerated, evidently paid little attention to publication; but Aristotle clearly recognised the value of the «book» for recording a systematic treatise and a detailed argument that requires to be studied at length, with the possibility of continuous reference back and forth, fro one part to another. It is fortunate that he did so, for the masters of the Graeco-Roman civilisation had not fo mulated their ideas in this way, there could have been n great revival of European culture at the end of the Middl Ages — the Renaissance.

The invention of printing from movalble type, whic made possible the production of large edition at relatively low cost, gave a tremendous impetus to the spread o knowledge, and established the role of the publishing industry as the base on which scholarship could bo founded The pioneers, such as Caxton in England and Aldus Manutius in Italy, were scholars as well as craftsmen, and often combined in themselves all the arts of editor, printer publisher and bookseller. Groups of scholars, whethe informal such as the Wandering Scholars, or formally constituted in «academies» like the Brethren of Sincerity a Basra, or the Academia Secretorum Naturae in Italy, ha not played any outstanding part in publishing, but ha not neglected to use the books as an instrument for spreading their ideas. This aspect of bibliographical histor has been well documented by Thornton and Tully [1].

The book, however, was not the only medium of communication. The letter formed an equally important medium, as it still does, but of a less formal character. This importance was recognised, however, and volumes of correspondence have always provided vital evidence of a first-hand nature for historians in all fields. The particular character of the letter, however, differed from that of the book. It was not, generally, used as a medium of expressing a full sequence of ideas which amounted to a complete picture of a given field; it was, rather, the means for communicating quickly and easily a single item of information to one who was certain to be interested, and thus it was, as a form, the direct ancestor of the modern periodical. It was the Scientific Revolution of the 16th and 17th centuries that brought together the formal group of like-minded scholars who converted into an institutional form what had previously existed as a rather haphazard enterprise between two or three friends.

The particular social need to be met was that of transferring highly specific knowledge so that it could be put to use; the motive pover came from the fact that these scholars were scientists, engaged in practical experimental work. They realised all too cleary how much they stood to gain by exchange of knowledge. Although they were naturally anxious to have their name linked with great discoveries, they also understood, in the words of the man who was one of the chief founders of the movement, Francis Bacon, that every man is a debtor to his profession. Furthermore, these men were gifted amateurs, in the sense that they were inspired more by love of knowledge than by hope of gain, and they were interested in the whole of knowledge as well as the particular field that they were themselves cultivating.

As is well known, the various trends in research and communication culminated in the foundation in England of the Royal Society, model and forerunner of many others. Its membership uncluded artists and men of letters as well as scientists; and they were encouraged to write about their work, not in a high-flown esoteric language, but in plain and simple English that corresponded to natural speech. The same qualities were to be seen later in one of the great founders of Russian science, M. V. Lomonosov, of whose style Pushkin wrote that it «springs mainly from his deep knowledge of the literary Slavonic language, and from a happy fusion of the latter with the language of the common people.»

The Royal Society was not the first periodical publisher, even in England. This had begun before the Civil War, in the field of political news and comment; but the Royal Society's «Philosophical Transactions» was the first scientific periodical to have a continued existence down to the present day, and T. H. Huxley once said that «If all the books in the world, except the «Philosophical Transactions», were destroyed, it is safe to say that the foundations of physical science would remain unshaken, and that the vast intellectual progress of the last two centuries would be largely, though incompletely, recorded.» Even in the nineteenth century, however, the «Philosophical Transactions» had, inevitably, to concentrate, after the Industrial Revolution, on the special fields of the physical sciences; the same vast progress had brought with it a large number of other such societies, representing not only research scientists and scholars, but also practitioners in the professional fields, law, medicine, engineering, theology, and so on. Commercial publishers, who had handled most of the book trade since the invention of printing, moved into the periodical field, mainly with literary and political journals, but also in specialist fields, of scholarship: perhaps the outstanding example is «Nature», founded in 1869 and still published by the house of Macmillan. This was begun as a deliberate attempt to reach a wider market for scientific reporting, and such a collaboration of scholars and business men, for the purpose of disseminating information, has often been repeated since.

The third great intellectual and industrial upheaval, sometimes called the Cybernetic Revolution, has intensified all the problems of formal communication methods. Many factors have contributed to this crisis: the great increase in numbers of research workers, who naturally want to publish; the consequent growth in size of the publishing industry; the spectacular changes in transport, which means many more opportunities for scholars to meet and discuss with each other;the consequent increase in the tempo of research and its extension into practice. None of these factors are new, of course, but the size of their growth has led to a radically new position, in which the old ways which had given satisfaction for so long have been subjected to new and intensive scrutiny. In particular, the whole of the communication process has been studied in the light of General Systems Theory, and the techniques of systems analysis, which has been developed as a direct result of the necessity for a detailed breakdown of any process that is to benifit from the use of computers.

The main principle of General Systems Theory is that of the inexorable interconnections between phenomena at all levels: this is the nature of the physical world, and although Man is certainly able to isolate entities and phenomena in order to examine them, this does not represent the true state of affairs. If conclusions are drawn solely on the basis of the static, the unconnected state, they may very well not apply in the dynamic interconnected state, which is the actual state with which we have to deal in reality. Our knowledge, so far as it goes — for it is, and will always be, incomplete, derives from our study and understanding of the real world; that world does not derive its phenomena from the processes of the human mind, though they may obviously affect it. This means that, while we may detach certain entities, or even groups of entities, from their natural dynamic state, is a mistake to suppose that our examination of them, in such a detached state, is complete. In systems analysis, it is a mistake to imagine that such a thing as a «closed» system exists in Nature, and this mistake is all too frequently made. For example, much of the contemporary discussion about library and documentation systems is either sterile, or grossly repetitive, because it fails, in the end, to regard these systems as part of a larger social system, and instead regards them as ends in themselves.

This is particularly marked when the question of automation arises; from some papers one would imagine that the purpose of an information service was to provide work for computers, instead of the function of computers being to facilitate information service to real people. This attiude seems to mar the otherwise useful conribution of T. Kitagawa to the February 1970 Meeting of FID/RI [2]. It is surely something of a limited view to see the main purpose of continuous education for Man in society, to qualify him, «to work with and to manage machines and systems in the cybernetical era». In similar vein, the Report of the Committee on Scientific and Technical Communication of the U. S. National Academy of Sciences — National Academy of Engineering speaks of the need for finding «ways of overcoming user apathy or resistance.» [3]. This is the language of the advertiser seeking to persuade consumers to buy products for which they feel no real need; whereas, as any librarian or information officer knows, users are eager to come by information that fills a gap in their knowledge — they do not have to be persuaded. If there is a reluctance on the part of users, the fault lies not in their stupidity, but in the nature of the goods they are offered. It is the «reprocessing and repackaging», in the SATCOM Report's [3] revealing phrase, that stands condemned, and the correct solution to the problem is not to expend effort to pile yet more paper on to scholars who are already overburdened, but to examine the whole process as a unified system in itself, to discover how the purpose of information transfer may be realised.

An early attempt to streamline the operation was the proposal advanced by J. D. Bernal at the Society's Scientific Information Conference in 1948. He envisaged, if not the cessation, at least a drastic reduction of periodical publication, and the deposit instead of individual papers at a central distributing agency. He withdrew this proposal in the face of fierce opposition from journal editors, but a significant reminder was given at the 1955 IUPAC Conference on the Documentation of Pure and Applied Chemistry (which has been lost to sight because it was never published as a single volume):

«J. D. Bernal ... already in 1948 ... proposed that papers should be submitted to a central agency who should classify and disseminate them to those who asked for papers on individual topics.

Bernal's proposal was certainly premature, but it is not too early now to begin thinking of the position as it will be, say, in 10—20 years hence. When that time comes journals will be twice their present size, relatively few Fellows will pay for them, and industry and the State will consider whether their subsidies which have become burdensome are really necessary; whether some better means of recording and distributing scientific results cannot be found».

This prophetic comment was made by R. S. Cahn, then editor of the «Journal of the Chemical Society», who had as good reason as anyone to see the approaching crisis. A further, much more detailed, memorandum was submitted in 1961 to the FID by I. Polzovics, who has repeated his warning in the VINITI volume, «International Forum on Informatics» [4], complaining that meanwhile nothing has been done: «Because of its significance, it seems reasonable to ring an alarm that wakes up this Sleeping Beauty from its slumber». Many other studies in these two volumes bear on the problem, from all aspects; that of M. J. Menou [5], in particular, emphasises the necessity for an overall study from the point of view of the individual scholar, who is both producer and user, and therefore marks both the beginning and the end of the knowledge transfer chain, or system [6].

In studying this system in detail, then, we have to start with the individual. His purpose, put simply, is to increase his control over his own environment so that he can manipulate it for his own ends. This requires understanding of how the environment works: as we said, it is not the creation of his mind, but exists independently of his mind. Control has to be acquired; it is not innate. Understanding, as Jean Piaget [7] has shown in his long series of experiments with children, comes primarily from activity, from interfering with the environment, from experimentation. This is how the scientist also works. But once the mind has developed, through the processes of seriation and classification, to the point of being able to work with abstractions (symbols) which stand instead of real phenomena, we can learn from records and benefit from other people's recorded knowledge just as productively as if we had discovered it for ourselves. Since the extent of these records is now so vast, the research worker clearly needs some agency to scan the general output and direct towards him those items that deal with the field in which he is working. When Francis Bacon took all knowledge for his province, he was presumably able to perform this function for himself, but the need for intermediaries became felt at least as early as the late seventeenth century, when the «Weekly Memorials for the Ingeniuous» were first published in 1682, and the Royal Society reported notes about the work of the Ingenious in all parts of the world.

We are now speaking, in fact, of the individual who wishes to be in communication with others in discussing a special field. If he knew all of them, presumably correspondence would be a direct, if rather tiresome channel. To this day, we use the channel for our intimate friends or for those far away whom we wish to interest in our work. But the attempt, two or three yers ago, to institutionalise this informal relation by Exchange Groups came to nothing, mainly because of the hostility of professional communicators such as editors and secretaries of learned societies. These are the organisers of knowledge transfer, the former by their journals, the latter by their meetings, conferences, symposia and so on. They are necessary, of course, because the scholar wants his work known not only to his friends, but also to the host of unknown «colleagues», to whom he can be linked by means of their common bonds — membership of the same societies, subscription to the same journals.

This is also a two-way relation, since the scholar as user, as receiver of knowledge, also profits by the network, and learns about interesting work being done by people unknown to him. This aspect is particularly important for younger scholars (who will be the leaders of tomorrow); for although it may still be true that members of the «invisible college», as D. J. de Solla Price has called the inner, senior membership of specialist groups, know of each others' work without the need for an information network, this does not apply to those who are not yet members, those who do not sit on the impotant committees and thus meet regularly, with time to talk to each other. The workings of the leading group often extend, in fact, to the regular, formal channels also. Diana Crane's survey of sociological articles shows that the academic characteristics of the contributors and editors of three «prestige» journals were very similar; the inference is that editors and referees look for qualities similar to their own in articles submitted for their consideration [8]. This applied even when the articles were anonymous, so that this similarity does not depend on the editors' personal knowledge of the authors.

The question of assessment is crucial. When an author decides to make his work public, he desires to submit it to informed scrutiny by experts in his field. The very fact of publication should be a guarantee that the work has reached a certain standard — not only in respect of contents, but also for presentation, due citation of previous work, and so on. There is a serious danger of breakdown here. The remarkable proliferation of new journals could be taken as a sign of healthy growth in the world of learning, but unfortunately it can also be due to over-commercialisation of the product, «information». The symptons are: lowering of standards, (usually due to absence of refereeing), duplication (the same article published in more than one place), excessive publishing of «news» (either of little interest, or duplicated), proliferation of commonplace «Notes» or «Letters to the Editor», and, above all, by misuse of the practice of commissioning article from well-known authors who very often have nothing new to say, or no new synthesis to attempt. The question is, should publication standards be set by market place, or by the objective assessment of informed specialists? There are dangers and advantages in both; it has been suggested by R. D. Whitley [9] that there is evidence that editors tend to favour articles that support the exising paradigm (in T. S. Kuhn's sense [10]), and are wary of publishing those that criticise. This could of course mean that the bias is against the genuinely novel, and one can indeed find instances to justify this. On the other hand, a proposal to depart from the paradigm should have to carry conviction; it must be more than usually well supported with evidence.

Reliance on the market now carries the stamp of SATCOM approval, because the survival of «for-profit» organisations depends on «their ability to recognize, understand and adequately serve users, and because of the management and marketing capabilities they have developed. . .»; but, as with all commodity production, there is nowadays an obvious tendency to rely more on the marketing capability than on the satisfaction of a real need, or the quality of the product. The philosophy of «built-in obsolescence» has entered the communications industry, and we read of lengthy surveys and complex calculations of the «half-life» of specialist journals. So long as libraries (in advanced countries at least) are well-endowed, journal publishers are assured of the minimum market necessary to make a profit, and the essential element of risk, if the SATCOM assessment is to be justified, is absent. Standards inevitably suffer, because factors ensuring ease of production take precedence over factors ensuring quality of product. There is therefore much to be said for Polzovics' view that all specialist journal publishing should be in the hands of learned and professional societies.

First publication of an author's work, whether as a book, pamphlet, periodical or report, introduces him to the social group whose connecting bond is formed by their association with work in a specialised field and their real need to keep informed of the work of others, known or unknown. The editor's or publisher's role is that of sponsor; he performs the introduction, implicitly guaranteeing that the work is both respectable and interesting. At the same time, he should be objective in his assessments, and equally guarantee to the author his right to publish his work — perhaps the most important argument in favour of control by the peer-group, the learned society. This is not the final judgement; that is reached by the decision of the «general public», if one may use that term of a special field. The work achieves, or does not achieve, consensibility, in Ziman's phrase; if it does, it becomes «public knowledge» [11].

At this point there enters into the network a new element: the librarian, documentalist, information scientist, or whatever name may be attached to those who have the social responsibility, not of printing and publishing original work, but of organising and putting to use the results of the publication process. «Publication» is used here in its broadest sense: that once the report of the work has been «uttered», or given out, it comes into the public domain and has to submit to an organising process so that its true purpose — that of communicating knowledge — can be realised. These «Messengers of Light», as Francis Bacon called them, bought to his New Atlantis «the books, and abstracts, and patterns of experiments of all other parts», and today they play an essential part in the communication network, even though their value may still, in some quarters, be underestimated.

It is not necessary to rehearse here either the functions of information workers, or the structure of the organisations which they serve; this has been done many times. «The International Forum on Informatics» contains a number of case studies, and A. Merta has applied systems analysis to the international level of operation [12]. The social role of all the members of this group is the same — to acquire, arrange and disseminate the knowledge contained in records that have come into the public domain. In this process, secondary records need to be created: abstracts, indexes, reviews. The bibliographical apparatus of a subject acts as a map of the territory; it is not the territory itself, but consists of a series of points of reference, which a user may study even though he is not on the ground itself. We may plot a route on a map in our own homes, and we may compile a list of references in a library, rather than in the laboratory. These references bring a user into contact with an author, even though he may not be directly linked to that author by any of the formal linkage systems already described.

The crucial importance of this function becomes clear if we remember that, unless an author's work becomes enmeshed into this secondary recording apparatus, it will be lost from sight as soon as the immediate circle of his acquaintance cease contact with the field of study. Communication with other workers who are unknown to an author, whether separated by space or time, is impossible without the aid of bibliography in one form or another.

How, then, should this type of «secondary recording system» develop? Technically, there is only one way open: we must make the best possible use of machines designed to handle large quantities of data, but this must be in the light of users' requirements, not the convenience of computer manufacturers. The futility of much so-called research in this area is demonstrated by the fact that we still do not know how to use the data produced by one of the most imaginative schemes yet designed to enlist the aid of computers — the MARC Project. Other schemes which have cost a great deal of money to develop and a great deal of publicity to sustain, often fall some way short of their ostensible target. Much more thought needs to be given to the psychological, linguistic and sociological foundations of Informatics, before we hand over the specification to the engineers.

There are hopeful signs that, on the sociological level, the requisite action has begun. The success of VINITI has encouraged other countries in Eastern Europe to establish central offices for scientific and technical information, some of which act as initiators of bibliographical systems, some as co-ordinators. The U. S. National Academy of Sciences and National Science Foundation have for long urged that a proper study of organisational problems be made, and their efforts seem to be bearing fruit through SATCOM — though it is not without significance that the membership of this Committee is heavily weighted in favour of the publishers and engineers: there are few scientists, no social scientists, no librarians, one Dean of a library school. They do, however, recognise the need for co-operation and co-ordination as well as competition, and the case for special grants for new services in libraries. One of the encouraging features of the Report is its view of the field as a whole, seeing documentation as an integral part of the wider area of scientific research and communication.

In practice, international co-operation is already under way in the fields of medicine, physics and chemistry, where learned bodies have taken the lead. They have already shown the advantage of this procedure, in their sensitivity to criticism and willingness to modify their schemes to meet the informed criticism of users. For future planning of international effort, the ICSU/Unesco project UNISIST represents an important attempt to find a central point, or hub, from which an integrated systems linkage may be managed. The organising committee of UNISIST comprised representatives of all, or nearly all aspects of the process of communicating knowledge. Organisationally, its range is excellent; but it seems to have been forgotten, for the moment at least, that just as no man is an island, so no subject exists in isolation either. Where does one draw the line between «science», and «social science», and «humanities»? Once again, we have to remember that, although distinctions can be made conceptually, these cannot be maintained successfully and fruitfully when it comes to the acid test of practice. There are many problems of the greatest difficulty facing the scheme, however, and if the necessary system of controls, channels, of information, facilities for initiating action where required, is to be constructed, it may perhaps be as well to start with an achievable target in view—always provided that the overall objective is kept in the longerterm prospect.

For this is nothing less than entry into a new phase in the history of the communication of knowledge; when all the lessons of the past have been learned and understood, and the aims and aspirations of those who produce and those who use knowledge have been taken into account; when modern technological achievements have been mastered and brought into the service of Man, instead of the other way about; when librarians and information officers see their role as actively facilitating the transfer of knowledge, as well as guarding the records; then we shall have designed and built an organisation to co-ordinate social effort for the achievement of a socially valuable purpose.

REFERENCES

1. J. L. Thornton, «Medical books, libraries and collectors». London, Library Association, 1949.
J. L. Thornton and R. I. J. Tully, «Scientific books, libraries collectors», 2nd ed. London, Library Association, 1962.

2. T. Kitagawa, Information science approaches to scientific information systems and their implications to sciences research In: «Proceedings of the Study Committee of FID/RI, Moscow 24—26 February 1970». Moscow, VINITI, 1970. pp. 54—80.
3. National Academy of Sciences — National Academy of Engineering. Committee on Scientific and Technical Communication. «Scientific and Technical Communication: a pressing national problem and recommendations for its solution». Washington, National Academy of Sciences, 1969.
4. I. Polzovics, Considerations to improve publishing methods of primary scientific and technical literature — especially periodicals. In: «International Forum on Informatics». Moscow, VINITI, 1969. Vol. II. pp. 464—478.
5. M. J. Menou, Problematics of the international systems for the transfer of scientific and technical knowledge. In: «International Forum on Informatics», Moscow, VINITI, 1969. Vol. I. pp. 217—238.
6. Seealso: A. I. Mikhailov and R. S. Gilyarevski, «Guide for an Introductory course in Informatics/Documentation». Paris, Unesco, 1970; especially the first three chapters.
7. Piaget's views are well summarised in his recent work «Genetic Epistemology». New York, Columbia University Press, 1970.
8. D. Crane, The gatekeepers of science: some factors affecting the selection of articles for scientific journals. «The American Sociologist», 1967, 2, 11, 195—201.
9. R. D. Whitley, The formal communication system of science: a study of the organisation of British social science journals. «The Sociological Review Monograph», No. 16, September 1970. pp. 163—179.
10. T. S. Kuhn, «The structure of scientific revolutions». Chicago, University of Chicago Press, 1962.
11. J. Ziman, «Public knowledge: the social dimension of science». Cambridge, Cambridge University Press, 1968.
12. A. Merta, A systems approach to the problem of building up international information institutions. In: «Proceedings of the Study Committee of FID/RI», Moscow, VINITI, 1970, pp. 44—52.

B. C. Brookes
(United Kingdom)

THE AGING OF SCIENTIFIC LITERATURE

1. Scientific Literature as a Continuous Renewal

In recent years emphasis has been laid on the exponential growth of scientific literature and the problems that follow from the so-called «information explosion». These problems look specially alarming to those who regard the development of science as a purely cumulative process, as an endless collection of new facts or data. But if such a view were valid all scientific papers would be of constant scientific interest whatever their date of publication, and clearly they are not. The development of science is more correctly to be regarded as a process of continuous renewal rather than of continuous cumulation. The literature of the recent past is constantly reviewed and updated; data already known are merged with new data and are redescribed and reinterpreted in the terms of newer theories. Working scientists concentrate their interest on the most recent publications and their contributions to the literature are to be regarded not simply as additions to a growing file but as corrections and refinements of papers published at some earlier period and which are thus superseded. If all the literature of science aged ten years or more were accidentally destroyed the historical record of a great human achievement would be lost but there would be little loss to science itself.

Though practical problems which face information practitioners today stem partly from the increasing growth of new publications, they also arise from the speeding up of modes of communication and transport. Scientists attend conferences on a world-wide scale and in large numbers and so keep themselves informed of trends and developments by direct personal contact. The tempo of scientific work has increased as a result and the increasing urgency of scientific activity has led to pressure to accelerate the distribution and dissemination of new scientific information especially when it can be applied to technological processes. Computerized systems have aided this acceleration. In the recent past, emphasis has been given to the increasing rate of growth of scientific publication partly because the evidence of this growth presents itself unambiguously and dramatically in the form of new periodicals to be processed. But the evidence of decay is less immediately discernible. As periodicals age, scientific interest in them gradually fades but the aging periodicals emit no discernible signal as they quietly pass from active scientific interest into the archives of the historian of science.

Scientific information centres therefore tend to carry a greater volume of obsolescent and obsolete material than is necessary. The greater the volume of obsolete material that has to be serviced and searched, the slower and more costly all information services become. Scientists have improved their chances of getting new information quickly by changing from dependence on libraries of the traditional form to dependence on computer-aided information centres which, ignoring the literature of the past, began their operational life by processing only newly published literature. The new computerized systems thus avoided current problems of obsolescence, but only for a time. As the databank of any scientific information centre ages, so the scientific value of its older files declines: magnetic patterns on tape or disc are no more immune from the decline of scientific interest than are marks on paper. Computerized systems have not yet faced the problems of weeding from their files the material that has already been superseded, but they will soon have to.

One way of reducing the effects of increasing growth is to learn how to recognise and then discard any material that has become of negligible scientific interest so that all the material retained and serviced by the working library or information centre remains of significant interest. The material ripe for discard is disposed of only if a national or central archival repository of scientific literature can be relied on promptly to satisfy any reference to material that has already been discarded by the working library, otherwise the discarded material can only be removed from the most active area to some other area where it can still be accessible. But, once a «bank» has been established, such as the National Lending Library for Science and Technology in Britain, the rationalization of obsolescent material becomes possible in all libraries and information centres which have access to it. The major problem of obsolescence is that though interest in the literature of science continues to decline as the literature continues to age, theoretically at least the interest can never be said to reach zero. So arbitrary decisions about discarding material have to be made and a rationale is therefore needed.

The volume of literature of active interest can be imagined as a wave-front moving through time (Fig. 1). New literature is generated continuously at an ever-increasing rate so the wave-front steepens and reaches a higher and higher maximum. But the new literature also begins to age, rapidly at first and then more and more slowly, its trace backwards in time being theoretically endless.

This present paper is concerned with techniques for measuring obsolescence. They have been devised to reduce the work of computation to a minimum. As the rate of obsolescence is a function of both the literature and also of its users' interests it is important that librarians know how to measure obsolescence rather than accept measures derived from other contexts and to understand the limitations of any measure they use.

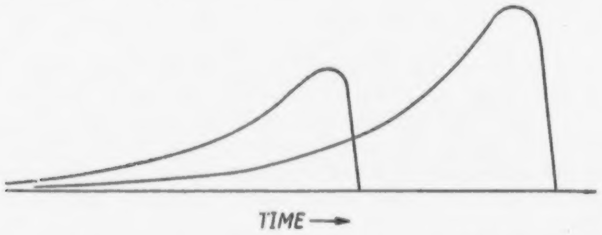


Fig. 1. The volume of literature of active scientific interest in an expanding science.

2. Analysis of a Typical Sample of Data

In order to provide numerical data with which to illustrate the practical technique of measuring rates of obsolescence, a sample of citations was taken from papers published in «Philosophical Magazine», vols. 19 and 20 (1969). All these papers related to the same subject — solid-state physics. All the citations noted were to papers already published so that references to papers «to be published» or «in the press» were ignored. «Private communications» even if dated were also ignored. The citations noted were mainly of papers published in periodicals but citations to reports and monographs were also included.

The citations were noted in three separate sets: all citations in (a) the January and February issues, (b) the May and June issues and (c) the September and October issues. These three sets of citations are displayed as frequency distributions in columns (1), (2) and (3) of Table 1 with the corresponding dates in years. The dates of column (4) are the sums of the three samples taken along the rows. The data of column (5) are derived from column (4) by summing the citations cumulatively from the earliest dates at the bottom of the column.

The first point to note about the three sample sets is that, in all three cases, the figures reach their maxima in 1967, illustrating that there is an inescapable delay between the use of newly published literature and the indication of this used by the appearance of citations to it. Having risen rapidly to their maxima the figures thereafter fall, in general, relatively slowly though with marked fluctuations from set to set in some years. It can also be seen that the years 1941—47 were unproductive of citations as compared with the years immediately preceding that period. This low productivity is less easily accounted for by sampling fluctuations because, though the figures are all very small, they persist in all three sets over the period. Clearly this period of low productivity can be explained by the fact that, allowing for delay between use and published citation, this period coincides with the period of World War II.

The total sample was split into three sets in order to illustrate the fluctuations that can occur when the negative exponential (or geometric) distribution is sampled. The exponential distribution has a thick short head and a long thin tail which reaches, theoretically, to infinity.

Table 1

A Sample of Citation Date (Phil. Mag. 1969)

Date	(1)	(2)	(3)	(4)	(5)
1969	—	15	24	39	1805
1968	47	86	71	204	1766
1967	65	87	96	248	1562
1966	76	70	63	209	1814
1965	49	51	53	153	1105
1964	33	50	67	150	952
1963	54	42	51	147	802
1962	43	29	30	102	655
1961	37	28	20	85	553
1960	18	11	21	50	468
1959	33	14	20	67	418
1958	14	13	21	48	351
1957	8	15	13	36	303
1956	11	9	18	38	267
1955	14	4	11	29	229
1954	7	4	7	18	200
1953	7	7	7	21	182
1952	8	7	10	25	161
1951	12	4	10	26	136
1950	4	5	12	21	110
1949	7	3	6	16	89
1948	4	2	6	12	73
1947	1	1	1	3	61
1946	—	1	1	2	58
1945	—	2	—	2	56
1944	1	—	—	1	54
1943	—	—	—	—	53
1942	—	1	—	1	53
1941	1	2	—	3	52
1940	1	3	3	7	49
1939	1	1	2	4	42
1938	1	—	1	2	38
1937 or earlier	10	11	15	36	36
	567	578	660	1805	

The relatively large errors that can arise from samples of citation (or usage) data have not been recognised hitherto in work on measuring obsolescence and are considered in more detail in Appendix II.

It is helpful always to split samples in the way described above. There is no logical necessity for citations to conform with the exponential or any other well known distribution. Though one would naturally expect to find a slow decline in numbers as the citation age, anomalies may well appear. If the sample has been split it is easy to check whether similar anomalies occur in all three sets or not. If the anomaly is found in only one set it is reasonable to dismiss it as a random fluctuation if no other cause is apparent. If the anomaly persists in all three sets, however, some causal explanation of the anomaly must be sought so that, if necessary, allowance for it can be made if the results are to be applied in practice. For example one of the sets may include citations from a long and important historical review while the others do not; such an anomaly emphasises the need to ensure that the total sample includes a fair proportion of the historical reviews that may occur in the particular literature being analysed. The data of Table 1, for example, remind us that normal scientific activity was seriously reduced by the last world war and that this is a factor which may sometimes affect the choice of date at which back runs of periodicals can safely be discarded.

In Fig. 2(a) the data of column (5) of Table 1 are plotted on linear graph paper and in Fig. 2(b) they are plotted in 3-cycle semilogarithmic graph paper which automatically takes logarithms of the numbers of citations but allows the time-scale to remain linear. The cumulated data of column (5) produce smoother graphs than the raw data of column (4). If the frequency distribution is exactly exponential or geometric then the graph plotted on semi-logarithmic should be a straight line. It can be seen that the graph is approximately linear from 1967 to 1947, that it levels out over the war period and then resumes its decline once more. By the time the graph has reached 1947 only about $3^1/_2$% of the total number of citations remain so that, for all practical purposes, only the initial slope need be considered. The straight line imposed on the graph to fit the plotted points was drawn by eye.

The rate of obsolescence is reflected in the slope of the linearity. This slope can be determined by regression analysis but for most practical purposes it suffices to estimate the slope directly from the graph. A straight line

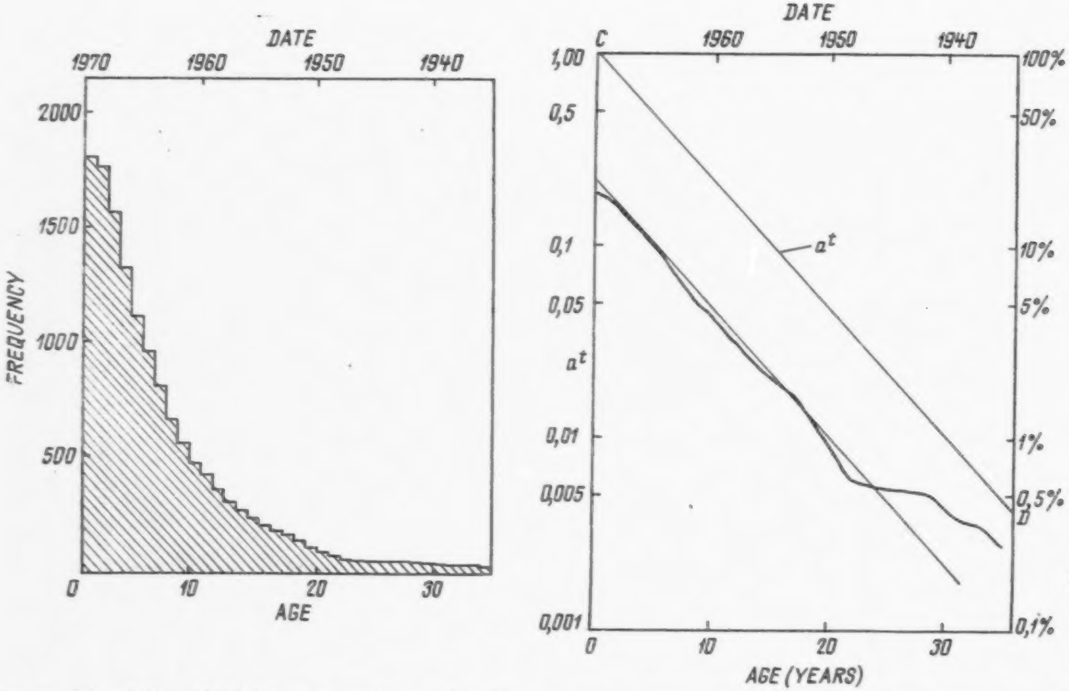


Fig. 2. Cumulative distribution of citations: Phil. Mag. 969 (a) linear scales, (b) semi-logarithmic, $(n = 1805)$.

CD is drawn very carefully from the top left-hand corner of the graph-paper parallel to the straight line of the graph. If the annual aging factor is **a**, the line CD is then the graph of $\mathbf{a}^{t}$ where **t** is the age in years. As $\mathbf{t}=1, 2, 3, 4, \ldots$ successively, the ordinates of the sequence of points correspond to $\mathbf{t}, \mathbf{t}^2, \mathbf{t}^3, \mathbf{t}^4 \ldots$ and their values can be read as fractions or percentages on the scale on the right-hand side. Any convenient point is selected though the higher the power chosen the more precise is the value of the aging factor derived from it.

From the graph it can, for example, be seen that $\mathbf{a}^8=$ $=0.295$. By succesive reference to square root tables we then have

$$\mathbf{a}^4=0.543,$$

$$\mathbf{a}^2=0.727,$$

$$\mathbf{a}=0.859\approx0.86.$$

This value of **a** can be verified directly from the graph by reading the value on the scale which corresponds to the point on CD where $\mathbf{t}=1$.

The fact that $\mathbf{a}=0.86$ implies that overall the number of citations in any specified year is 0.86 times the number of citations in the previous year throughout the practical life of the periodical.

Reference to the line CD can now save much calculation in answering questions relating to the obsolescence of this particular periodical. For example, a second measure of obsolescence that is frequently used is the half-life, **h**. This half-life is the age at which the usage of the periodical or set of periodicals has fallen to 50% of the total usage. It is therefore given by the age at which CD crosses the 50% level on the graph, i. e. 4.5 years approximately. A better estimate can be obtained by noting the age at which the total usage has fallen to 25%, a period which will correspond to 2 half-lives. In the example we have used, we find that CD crosses the 25% level at 9.0 years, thus confirming the first estimate.

The measures **a** and **h** are clearly related. They must satisfy the equation $\mathbf{a}^{\mathbf{h}}=1/2$ from which

$$\mathbf{h}\log\mathbf{a}=-\log 2$$

$$\text{or } \mathbf{h}=-\log 2/\log\mathbf{a}.$$

As $\mathbf{a}<1$, log $\mathbf{a}$ will also be negative so that $\mathbf{h}$ is positive. In most practical calculations it will be found that it is more convenient to use $\mathbf{a}$ than $\mathbf{h}$.

Further reference to the line CD shows, for example, that 90% of the citations will be less than 15 years old, that 95% will be less than 20 years old, and that if issues of the periodical were discarded after 12 years then 16% of the references would be lost.

3. The Concept of Utility

Consider the geometric series.

$S[0, \infty] = 1 + \mathbf{a} + \mathbf{a}^2 + \mathbf{a}^3 + \ldots$ to infinity (1). When $\mathbf{a}<1$, this series converges to the limit $1/(1-\mathbf{a})$.

The shortened series from which the first $\mathbf{t}$ terms are omitted is

$$\begin{aligned} S[t, \infty] &= \mathbf{a}^t + \mathbf{a}^{t+1} + \mathbf{a}^{t+2} + \mathbf{a}^{t+3} + \ldots \\ &= \mathbf{a}^t (1 + \mathbf{a} + \mathbf{a}^2 + \mathbf{a}^3 + \ldots \\ &= \mathbf{a}^t . \ S[0, \infty] \end{aligned}$$

and so converges to the limit $\mathbf{a}^t/(1-\mathbf{a})$.

Thus the series obtained when we put $\mathbf{t}=0, 1, 2, 3, \ldots$ successively in $S[\mathbf{t}, \infty]$ is also a geometric series with ratio a as in the series (1).

The data of column (4) of Table 1 (apart from two initial build-up terms) are approximately proportional to the terms of the series $S[0, \infty]$ whereas the data of column (5) of Table 1 are approximately proportional to the terms of the series $S[t, \infty]$ with $\mathbf{t}=0,1,2,3\ldots$ successively.

If the obsolescence of a selected literature produces a linearity when the cumulative sums are plotted as described in the previous section, then it implies that they form a geometric series with a ratio $\mathbf{a}$. We have noticed that in using citations there is a delay of about two years before present usage of a paper can be relied on to appear as a citation in a subsequent paper. It is therefore not unreasonable to ignore the first two points of the plotted graph and to base all calculations concerned with usage on the fitted straight line only. Thus all calculations are simplified with no serious loss of precision.

Clearly $S[t, \infty]$ is proportional to the total number of citations that can be expected to be made to the literature when it has attained the age of **t** years. On the assumption that numbers of citations are proportional to usage, then we would expect the total usage of a newly published issue of a periodical which has an annual aging factor of **a** to be $S[0, \infty]$ which equals $1/(1-\mathbf{a})$, i. e. that total usage of the periodical is limited and the limit can be predicted. The quantity $1/(1-\mathbf{a})$ is therefore important. It is called the **utility factor** and is denoted by **u.**

Clearly, the more rapidly a periodical ages the smaller are the numerical values of both **a** and **u**, i. e. the total usage decreases if the periodical ages more rapidly. As for solid-state physics we have estimated the annual aging factor to be 0.86, the utility factor is $1/(1-0.86)=1/0.14=$ $=7.14$. This result implies that the total usage or utility of de complete output of a periodical literature on solid-state physics is equivalent to only 7.14 times the utility of one year's output — a result which emphasises the ephemeral nature of scientific periodical literature. It is a measure of the volume of the wave depicted in Fig. 1.

In practice, in any particular special library or information centre, we are normally concerned with relative rather than with absolute measures of use. The special library will be acquiring a number of periodicals which all contribute papers to the scientific subject of interest. But some periodicals will carry more relevant papers than others and so will have a higher initial utility if we give equal weight to all relevant papers. It is reasonable to assume that all the **papers** relevant to a given subject age at the same rate but as some of the periodicals carry papers on other subjects also it cannot be assumed that all the periodicals age at the same uniform rate because papers on different subjects may age at different rates. But as the library will be concerned only with relevant **papers**, the rate of aging of the periodicals within that library will be determined by the aging of only the relevant papers they contribute.

If the two periodicals $\mathbf{P}_1$ and $\mathbf{P}_2$ contribute $\mathbf{N}_1$ and $\mathbf{N}_2$ relevant paper per annum respectively, then the residual utilities of $\mathbf{P}_1$ and $\mathbf{P}_2$ of the **t** years will be

$$\mathbf{U}[\mathbf{P}_1, \mathbf{t}]=\mathbf{k}\mathrm{N}_1\mathbf{u}\mathbf{a}^{\mathbf{t}}$$

and

$$\mathbf{U[P_2, t] = kN_2ua^t}$$

where **k** is a constant. We therefore have

$$\mathbf{U[P_1, t]/U[P_2, t] = N_1/N_2}$$

for all values of **t**, i. e. the residual utilities always remain in the same ratio as the unitial utilities of the periodicals.

4. Synchronous and Diachronous Perspectives

Sample counts of citation, usage or reference have to be taken over a relatively short period, otherwise the results will be blurred. At the time they are taken such data provide estimates of the intensity of interest in the particular literature, distributed over all back issues of the periodicals, as they stand at the time of the count. But does such a view of the literature as it is now enable us to predict the obsolescence patterns of, say, 1981? Do the data we have been analysing, which include data for the year 1959 and earlier, reflect the pattern of obsolescence as it would have appeared in 1959? Are rates of obsolescence changing? Have they been affected by the operation of computer-aided S. D. I. systems? Is a rate of obsolescence affected by the rate of growth of the literature? We need answers to such questions in order to establish the reliability of the projections that are needed for planning information systems rationally.

In considering such problems it is helpful to distinguish two ways of regarding a run of annual periodical volumes. The run can be regarded either:

(a) **Synchronously** — when we consider the utility of the run as a totality **now**, or

(b) **Diachronously** — when we trace the declining utility of some particular annual volume as it ages year by year from its date of publication.

If both the annual output of papers and their rate of obsolescence remained constant year by year, the diachronous tracing of the declining utility of some particular volume through the years would be exactly mirrored by the synchronous view of the declining utility of the actual volumes at present occupying the shelves. But most literatures of science are growing rapidly and it is not known whether the annual aging rate is affected by rate of growth. Clearly, this is an empirical problem; the answer can be obtained only by making suitable measurements.

For example, Fig. 3 shows the results of plotting a citation count (519 citations) on articles (relevant to solid

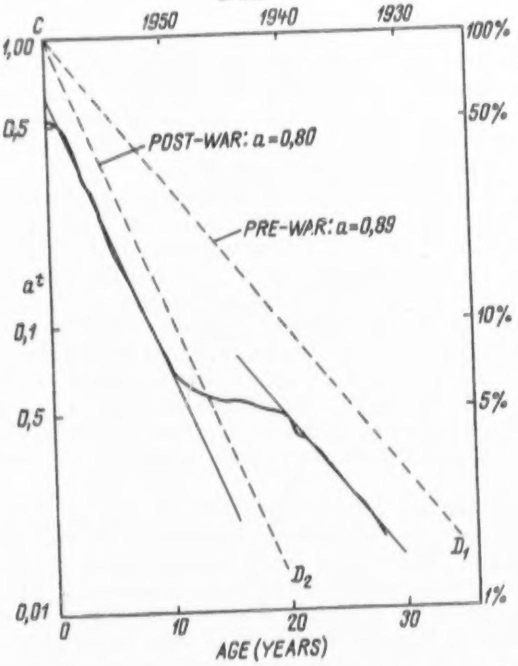


Fig. 3 Cumulative distribution of citations: Phil. Mag. 1959 solid-state physics only (n=519).

state physics only) in the 1959 volumes of «Philosophical Magazine». The plot shows two linearities, separated by the war period, of distinctly different slopes. The annual aging factor is 0.89 for the pre-war and 0.80 for the post-war linearity. The figure for 1969 has already been shown to be 0,86 (based on a sample of citations). Are these va-

FILMED FROM BEST AVAILABLE COPY

lues of a significantly different and, if so, can they be related to growth or to any other measurable factor?

Clearly, more measurements of more literature are needed to clarify these problems relating growth and obsolescence.

5. Simple and Composite Literatures

The literature of solid-state physics, which provided the data for the illustrative examples of the previous sections, conforms closely with the negative exponential law of obsolescence and yields linearities on semi-logarithmic graph paper as demonstrated. But not all scientific literatures conform with this simple law.

If a literature has two or more distinguishable components they could be the theoretical and the observational aspects of a science, the theoretical and the practical aspects of an engineering subject, or the theoretical and taxonomic aspects of a biological subject — it may be found that the two components age at different rates. This effect was noted by Meadows in his study of the obsolescence of the literature of astronomy in which he found an «immediate» interest in theoretical astronomy, which merges with theoretical and space physics, and a «historic» interest in descriptive astronomy which has a very long history.

Unfortunately for the analysis of aging, even if the aging of the two components each conforms precisely with the negative exponential law though at different rates, the aging of the combined literatures taken as a unity does **not** conform with a simple exponential law. The effect is illustrated by Fig. 4. In Fig. 4 (a) the two components P and Q of equal size, have annual aging factors of 0.97 and 0.81 respectively. In Fig. 4(b) the two components, of the same aging rates, 0.97 and 0.81 are in the ratio of 1 to 3. It can be seen that in both cases the composite result of adding the two linearities does not provide a third linearity. This means that this is no solution of the equati

$$qe^{-\alpha t} + (1 - q)e^{-\beta t} = e^{-\gamma t}$$

where q is a fraction and α, β, γ are constant aging factors, which is valid for all values of t, except for the trivi-

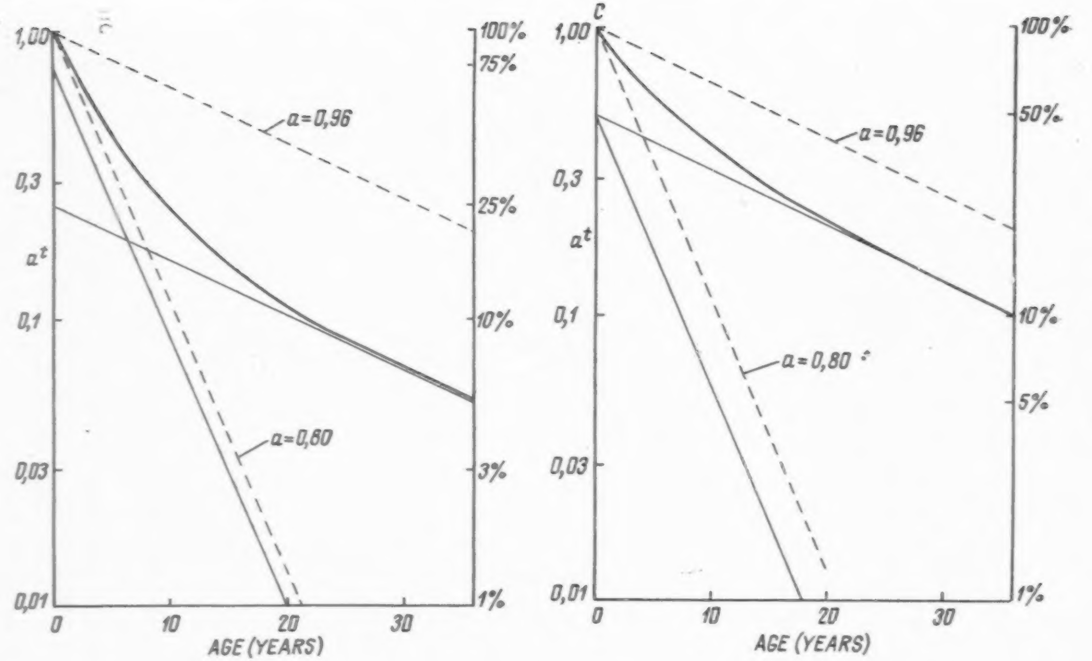


Fig. 4. Composite literautre with two components of different aging rates. (a) Components 1 : 1. (b) Components 3 : 1.

FILMED FROM BEST AVAILABLE COPY

al case in which $\alpha=\beta=\gamma$. In general it can be seen that initially the more rapid rate of decay is the predominating factor but that in the long run the slower rate predominates.

The problem of sorting out the components of a sample of citations from such composite literatures can be solved graphically, if need be, by taking a sample which is large enough to give a reliable «tail» of the older citations. The tangent to this tail is produced to run back to t=0 (Fig. 5). The ordinates of points on this tangent are noted and subtracted from the ordinates of the original plotted curve using, of course, the logarithmic rather than a linear scale. The reduced ordinates are then replotted. If the original composite curve has only two components,

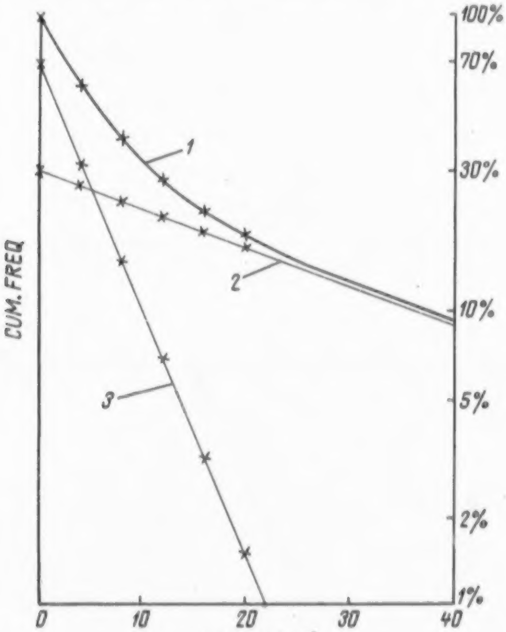


Fig. 5 Graphical analysis of a literature with two components of different aging rates

then the second plot **may** be linear; if so, then the two components will have been separated.

However, further difficulties are likely to be met. This graphical analysis would work only if the proportions of the two components remained constant over the period covered by the analysis. If **q** is a function of time also, i. e. if the proportions of the components are changing with time, as they may well be, then the above graphical analysis will not yield clear-cut solutions.

From a practical point of view also, any library or information centre which provides an information service for a composite subject cannot efficiently serve both its «immediate» and its «historic» users with a single set of periodicals cut off by discard at some compromise point. The «immediate» users could justifiably complain of the obsolete clutter in which their own papers were embedded and the «historic» users could justifiably complain that the compromise discard point was unfair to them. It would be preferable to separate the two components and to service them separately even though some of the current periodicals in the subject may be contributing to both components.

Subjects of some breadth, such as «chemistry» or «physics» or «medicine» are certainly composite, with several components. They are most unlikely to conform to simple exponential laws. This fact implies that measurements on obsolescence must be done on **papers** rather than on periodicals and with good bibliographical control.

In analysing the aging curves of composite subjects, the distinction between synchronous and diachronous perspectives must also be remembered. A composite curve gives **only a synchronous** perspective. It would be erroneous to interpret a curve as indicating that the rate of obsolescence of the subject has been changing, because such a curve may be the result of conflating two components which age at different rates. Studies of rates of change of obsolescence require diachronous data, e. g. estimates made directly on the literature published at intervals over the period of interest.

A general science library serving a number of scientific subjects and several different immediate, historic and other kinds of interest will be forced to compromise. It is the easier to do so if it is backed by an efficient regional or central 'bank' of scientific periodicals. It becomes increasingly important to the development of efficient and economic information services, all grappling with the problems of both growth and obsolescence, that hierarchical services with clearly defined ranges of literature should be established.

6. Citation and Usage

The data used in this paper were derived from citations in a specified periodical and quoted by papers which contribute to solid-state physics. Such citations are assumed to indicate the actual use made of the literature by the scientists who contribute to it. They have the advantage of being objective in the sense that the same citation data are observable by all who care to interest themselves in them.

But the exact relationship between the actual library usage of the literature and the use of citations is not known. If the library is serving users who are not all contributors, then citations are most unlikely to be indicators of actual usage. A university library, for example, provides for research scientists, students and historians of science and possibly for general readers also. In such cases the use of citations must be abandoned for records of actual usage within that particular library.

It will almost certainly then be found that the data do not conform with the geometric or any other standard distribution. If a measure of obsolescence is required, the mathematics and measures used in this paper do not apply. The only solution is to express the cumulative data as percentages of the total, to draw the graph of these percentages (log scale) against time (as before) and then make all estimates directly from the graph and without the use of any measures derived from the geometric distribution.

7. Obsolescence and Growth

Is the rate of obsolescence of a scientific literature related to its rate of growth? This is an empirical question which can be answered only by measuring the rate of obsolescence of a growing literature at intervals. But, as differences are likely to be small, the measures must be based on very large samples to ensure that any differences found are statistically significant. It would therefore be helpful to consider the question theoretically to see what might be expected and to formulate hypotheses to be listed.

If the rate of growth of the literature is measured in terms of the number of papers published per annum, it is reasonable to assume initially that the productivity of the contributing scientists remains constant, or approximately so, and that the growth of the number of papers published indicates that there is a concomitant growth in the number of contributors. It will also be assumed that, overall, the new contributors use the literature and, in particular, distribute their citations as the contributors whose ranks they join. This assumption implies that the **utility** of the literature increases at the same rate as the number of contributors.

If **g** and **s** are the annual growth factors of the number of papers and of the number of contributory scientists respectively, then

$$\mathbf{N(T)=N(O)g^{t}} \qquad (1)$$

and

$$\mathbf{U(T)=U(O)s^{t}} \qquad (2)$$

where **N(T)**, **N(O)** and **U(T)**, **U(O)** are the numbers of papers published and the utilities in the years **T** and **O.**

The utilities are also related to the corresponding annual aging factors, $\mathbf{a}_O$ and $\mathbf{a}_T$, by the relations

$$\mathbf{U(0)}=\frac{N(O)}{1-a_O} \qquad (3)$$

and

$$\mathbf{U(T)}=\frac{N(T)}{1-a_T} \qquad (4)$$

We now seek the relationship between $\mathbf{a}_T$ and $\mathbf{a}_O$. The terms **(U)T** and **U(O)** are eliminated from (2) by substituting the expressions (3) and (4) to give

$$\frac{N\,(T)}{1-a_T}=\frac{N\,(O)\,s^T}{1-a_O}$$

Then, from (1)

$$\frac{N\,(O)\,g^T}{1-a_T}=\frac{N\,(O)\,s^T}{1-a_O}$$

or

$$\frac{g^T}{1-a_T}=\frac{s^T}{1-a_T}$$

which yields

$$\frac{a_T-a_O}{1-a_O}=1-(g/s)^T \tag{5}$$

The expression on the right hand of (5) is positive if $\mathbf{s}>\mathbf{g}$, zero if $\mathbf{s}=\mathbf{g}$ and negative if $\mathbf{s}<\mathbf{g}$. So $\mathbf{a}_T$ is greater than, equal to or less than $\mathbf{a}_O$ if $\mathbf{s}$ is greater than, equal to or less than $\mathbf{g}$. At least these are the results that would be expected on the stated assumptions.

It can be seen therefore that any change in the value of the annual aging factor is likely to arise only if there is a change in the average productivity of the contributing scientists as measured by the numbers of papers they published per annum.

The measure of productivity in the direct sense would be difficult to estimate because there is no accessible record of contributing scientists who do not happen to publish at least one paper in the period of measurement. So, by various sampling techniques, estimates need to be made of $\mathbf{g}$, $\mathbf{s}$, $\mathbf{a}_O$ and $\mathbf{a}_T$ for some convenient value of $\mathbf{T}$.

Estimates of these quantities as they apply to solid state physics for the years 1963 and 1968 have been made by Oliver [2]. The results are: $\mathbf{g}=1.134$, $\mathbf{s}=1.150$, $\mathbf{a}_O=0.78$, $\mathbf{a}_T=79$. Her consideration of the errors arising from her sampling techniques leads her to conclude that the difference between the estimate of $\mathbf{g}$ and $\mathbf{s}$ is not significant at the 5% level. As the annual aging rates, based on citation counts, are equal, Oliver concludes that the above theoretical analysis is corroborated, i. e. that rates of obsolescence are not sensitive to rate of growth but are sensitive to changes of productivity.

Vlachy [3] has recently shown that changes are taking place in the degree of multiple authorship of scientific papers. It may therefore be necessary to consider how these changes effect measures of «productivity» and so a further factor is introduced into any analysis of the relationship between growth and obsolescence.

8. Applications

The main application of measures of obsolescence lies in the rationalization of the periodical stocks of special libraries and in the planning of information systems.

If in a special library the subject of scientific interest happens to be 'simple', such as solid-state physics, then we can assume that scientific interest in the periodical stock declines as the stock ages, in the case of solid-state physics at a rate of 14% per annum, in a way which spreads this rate of decline uniformly over the complete runs of all the contributing periodicals. Especially if there is a central library able to supply copies of early papers required, there seems to be little purpose in retaining back numbers indefinitely since they take up shelf-space and require servicing even if they are rarely used. But what should be the guiding policy for the librarian who wishes either to remove the little-used volumes to a back store or to discard them completely?

Hitherto, where discard has been practised after the rate of obsolescence has been measured, all periodicals have been discarded at the same age to leave equal runs (measured in years) on the shelves with immediate access. This technique is simple to apply but it takes no account of the fact that some periodicals are more useful than others. The result is that the most highly productive periodicals are discarded at an age when their residual utilities are still relatively high whilst the least productive periodicals are retained for longer than their useful life. This technique is therefore still wasteful of shelf—space and is clearly not the best solution to the problem.

The principle that is proposed relies on the concept of utility. Some acceptable minimum utility is decided upon and a periodical is discarded only when its residual utility has declined to this minimum level. The result of applying this principle is that the runs of retained periodicals are no longer equal in general; the most useful periodicals are retained much longer than the least useful. Instead of retaining equal runs, 'tails' of equal utilities are discarded.

If the minimum acceptable level of utility, i. e. the discarded utility, is $\mathbf{d}$, then the age of discard of any periodical with initial utility $\mathbf{i}$ is given by

$$\mathbf{a}^{t} = \frac{\text{discard utility}}{\text{initial utility}} = d/i \qquad (1)$$

If $\mathbf{a}$, $\mathbf{d}$ and $\mathbf{i}$ are known, the corresponding value of $\mathbf{t}$ can be read from the graph of $\mathbf{a}^{t}$.

The remaining problem is to quantify $\mathbf{d}$ and $\mathbf{i}$, though it should be noted that only the **ratio** of these quantities is needed. The simplest way is to consider the minimum level of acceptance of periodicals. In any real library some periodicals will have been selected for acquisition and others, considered for possible acquisition, will have been rejected because they offer too little that is relevant to the subject. It should therefore be possible to examine the most marginal of the periodicals acquired and to determine the average number of relevant papers, $\mathbf{n}_1$, it yields per annum. If some other periodical already acquired yields $\mathbf{n}_2$ relevant papers per annum ($\mathbf{n}_2 > \mathbf{n}_1$), we have

$$\mathbf{a}^{t} = \mathbf{d}/\mathbf{i} = \mathbf{n}_1/\mathbf{n}_2 \qquad (2)$$

as required. This method has been described in more detail elsewhere[4].

A second method of determining the required cutoff may be based on the cost of acquiring loans or photocopies of the papers called for from the tails that may be discarded. As estimate is needed of the total cost of acquiring a loan or photocopy, including an estimate of the cost of the clerical work involved and the cost to the borrower of the delay in acquiring the required copy externally as well as postage and any other direct charges. Let his total cost be $\mathbf{p}$ monetary units per photocopy or loan. The principle to be applied is that the periodicals are discarded when the cost of the substitute photocopies becomes less than the costs of retaining the periodicals on immediate access.

During the $(\mathbf{t}+1)$th year of the life of any periodical its utility will be $(1-\mathbf{a})\ \mathbf{a}^{t}$ or $\mathbf{a}^{t}/\mathbf{u}$. The number of demands made for papers within this particular annual volume will depend on two factors: first, the average number of relevant papers it contained initially — say **n** — and, secondly, the intensity of local usage, say **b**, of the particular periodical per annum. This factor **b**, which will vary from one library to another, can best be determined by noting the number of demands for it during any temporary absence of the periodical such as when, for example, it is sent for binding. (Because demands for a periodical decrease as it ages, some correction, which can be obtained from the graph of $\mathbf{a}^t$, may have to be made to the value of **b** derived from the count of demands). The total number of demands made for papers in the volume of age $(t+1)$ years will thus be $\frac{a^t nb}{u}$.

The total cost, **s**, of retaining the periodical on the shelves for one year, again allowing for its proper share of current servicing costs, is also needed. The average cost of a demand for one of the papers it contains will then be

$$\frac{su}{a^t nb} \qquad (2)$$

This will come at a time when the quantity (2), which increases as **t** increases, becomes greater than **p**. At equality we have

$$\frac{su}{a^t nb} = p$$

or

$$a^t = \frac{su}{pnb} \qquad (3)$$

and the value of **t** can be found by reference to the graph of $\mathbf{a}^t$.

Clearly this second method demands a greater detailed knowledge of the costing of the library services than is normally available.

APPENDIX I.

The negative exponential and geometric distributions. Because time is a continuous variable, obsolescence is a continuous process for the theoretical analysis of which the continuous negative exponential is the most appropriate distribution. The probability density function is

$$p(T) = \frac{1}{\alpha} e^{-T/\alpha}, \; 0 \leqslant T < \qquad (1)$$

where α is a constant. But in applying the formula to the aging of literature, we find that dates of citation are usually available only in terms of the year of publication. The continuous time variable is thus broken up into a sequence of discrete periods of one year. We thus count the number of citations to papers published in particular years and take the year as the unit of time.

In a distribution of citations which conform with (1) above, the probability, $\mathbf{P(t)}$, that a particular citation is earlier or older than $\mathbf{t}$ years is given by

$$\begin{aligned} p(t) &= \int_t^\infty \frac{1}{\alpha} e^{-T/\alpha} dT \\ &= e^{-t/\alpha} \end{aligned} \qquad (2)$$

If we put $\mathbf{e^{-1/\alpha} = a}$, we have

$$\mathbf{P(t) = a^t} \qquad (3)$$

which forms a geometric series as $\mathbf{t}$ takes the values 0, 1, 2, 3, ... in sequence. The probability of a citation other than $\mathbf{t}$ years but not older than $\mathbf{(t+1)}$ years will then be

$$\begin{aligned} \mathbf{P(t) - (P+1)} &= \mathbf{a^t - a^{t+1}} \\ &= \mathbf{a^t (1-a)} \end{aligned} \qquad (4)$$

which is the general term of the basic geometric series used in this paper.

The discrete geometric distributions (3) and (4) are thus used as convenient and close approximations to the negative exponential. A similar argument shows that in the analysis of growth a geometric series, with a factor greater than one and therefore divergent, is an equally convenient approximation for bibliographic analysis to continuous exponential growth.

APPENDIX II.

The sampling error of the geometric distribution. Measurement of the rate of obsolescence always involves the sampling of citations and sampling always involves possible errors however carefully the sampling work may be done. Any particular value of a measure of aging derived from a sample must be regarded not as **the** true value but merely as a single random sample of all the possible values that could be derived from other different samples of the same size taken independently from the same source.

The exponential or geometric distributions which underlie all obsolescence measures give rise to relatively large sampling errors which may account for some of the apparent lack of agreement between results reported by different workers on the same body of literature and which have led to some uncertainty in application. These sampling errors require brief discussion with particular reference to the sampling of the geometric distribution.

If the geometric distribution has the form $(1-\mathbf{a})\ (1+\mathbf{a}+\mathbf{a}^2+\ldots$ $\ldots\mathbf{a}^t+\ldots$ to infinity then the mean of this distribution is $\frac{a}{1-a}$ and its standard deviation, which is a measure of the spread or dispersion of a distribution is

a. So we write
1 — a

$$\text{mean, } \mu=\mathbf{a}/(1-\mathbf{a})$$

$$\text{s. d., } \sigma=\mathbf{a}/(1-\mathbf{a}).$$

Estimates of the mean derived from samples of size **n** would form a second distribution which has the same mean, μ, as the distribution which is sampled but will have the smaller deviation $\sigma/\sqrt{n}$ This second distribution of estimates of the mean will have a different form. If **n** is large enough, as will usually be the case in measuring obsolescence, the second distribution can be assumed to be normal or gaussian. This distribution is symmetrical about its mean, (Fig. 6).

The standard deviation of the destribution of estimates of the mean will therefore be $\sqrt{\mathbf{a}}/(1-\mathbf{a})\ \sqrt{n}$. It is known that 95% of all such estimates will be written $\pm 1.96\sigma$ of the mean value **u**. Hence 5% of the estimates will have en error greater than $\pm 1.96\sigma\sqrt{\mathbf{n}}$ which for the geometric distribution becomes ± 1.96 $\sqrt{\mathbf{a}}/(1-\mathbf{a})\sqrt{\mathbf{n}}$.

To express this approximately as a percentage error we divide the above expression by μ and multipyl it by 100. After reduction the required percentage is

$$\pm 196/\sqrt{an} \qquad (1)$$

If we wish the error to be less than 10% of μ (with a probability ot 95%), the critical sample size is given by $\frac{196}{\sqrt{\mathbf{an}}}=10$, from which

$$\mathbf{an}=19.6^2\approx 400.$$

As $\mathbf{a}<1$, $\mathbf{n}>400$, i. e. the minimum sample size for obsolescence measures with reasonable expectation of yielding an error of less than 10% in 400, a figure which increases to 500 if **a** is known to be 0.80, for example. Unfortunately, precision increases only as the square root of the sample size so that to reduce the possible error to 5% would require the sample size of 500 to be increased to 2000.

There is no escape from sampling errors whatever obsolescence measure may be selected and whatever technique may be used to derive its value. The errors are inherent in the sample.

Even the above discussion of errors is based on the assumption that the sample has been well randomized. It has also become obvi-

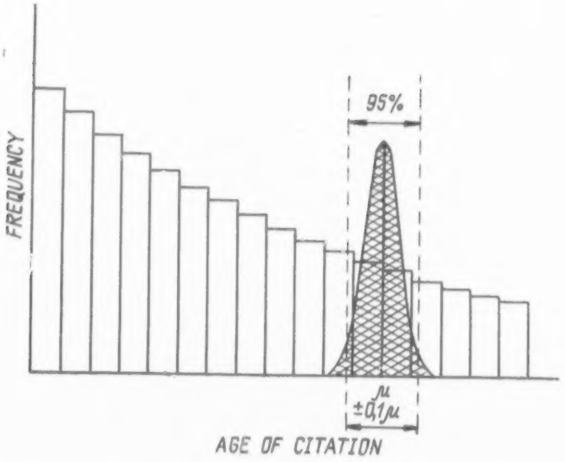


Fig. 6 The error distribution of the mean of a geometric series with $a=0.80$. $(n=400)$.

ous that the literature whose obsolescence is being measured must be well defined bibliographically so that others who may be interested can repeat the measure on their own sources and get comparable results.

In quoting the results of measurements on obsolescence it is therefore necessary always to quote the sample size and to specify the literature studies as precisely as possible.

REFERENCES

1. A. J. Meadows. The citation characteristics of astronomical literature. «J. Doc.», 1967, № 1, 28—33
2. M. R. Oliver. «J. Doc.», 1971, 27, № 1, 11—18. «The effect of growth on the obsolescence of semiconductor physics literature»
3. J. Vlachy. Physics periodical in retrospect and comparisons. «Czech. J. Phys.», 1970, B 20, p. 501—526
4. B. C. Brookes. The optimum p% library of scientific periodicals. In: Proceedings of the meeting of the Study Committee of FID «Research on the Theoretical Basis of Information», Moscow, 24—26 February 1970.
5. B. C. Brookes. Obsolescence of special library periodicals: sampling errors and utility contours». J. Am. Soc. Inf. Sci.», 1970, 21, № 5, 320—329.

9th October 1970

J. Wanke
(Czechoslovakia)

CONFLICTING PHENOMENA IN IR SYSTEMS

The art of Information Retrieval (IR) has reached a level where standard technologies and methodologies are being accepted. In discussing the theoretical aspects of this art, no distinction is generally made as to the different practical application branches of IR Systems. In our discussion, we shall treat the distinctive features of two extreme cases of IR Systems, viz. what we may call a «true» documentation IR System (say, a patent library), and a business data IR Systems (such as for inventory control, portfolio valuation, or flight reservation).

Essentially, there is no difference between searching a documentation file or a business data file, as both systems use much the same requisites and technologies. On the other hand, they may differ — and in most cases actually do so — in certain points, the importance of which has been somewhat neglected up to now as compared to the importance attributed to other system parameters (e. g. file size, access rate, relevance, noise, recall ratio, etc.).

The points considered here are: (a) file utilization, (b) mean record lifetime, (c) mean record length, and, (d) mean item value.

(a) First of all we shall try to define a measure of the utilization of the file, or of the «turnover» of the information stored. Let us relate the size of the file to the number of queries requested during a given time period by the simple formula

$$D = q/F \qquad (1),$$

where D is the daily duty factor, or measure of the file's utilization.

q is the everage number of queries per working day, and F is the size of the file (number of records, each record corresponding to one item).

In a more refined form which will be discussed later, the measure of the file's utilization will involve also the volume of an answer averaged over a high number of queries.

The (daily) duty factor D depends on the nature of the service. In a news agency with several million records in the file, about 600 simple searches (e. g. for biographical data, photographs etc.) are made daily which corresponds to D of the order of 10^{-4}. On the other hand, in a conventional UDC file in the Prague Central Technical Library with about the same number of records, some 2000 searches were made in 1960, i. e. less than 10 searches a day on the average. This corresponds to a D figure two orders of magnitude lower than in the previous case. In small peek-a-boo card files with a few thousand records a typical D figure might be somewhere in the 10^{-5} through 10^{-6} range. Generally speaking, in a given field of knowledge, D is to be expected to be invariable over large ranges of file sizes.

(b) Moreover, there is also a difference in the average life time of a record in the file. Thus while the average life of a record in a flight reservation system will not surpass a few weeks, an item in a documentation record file stays there for years.

Therefore the price of a bit (or byte, word ... etc.) stored is not determined only by the total storage price vs storage capacity ratio. A time factor corresponding to the mean life time of an item should be also considered. In the old days of paper card files, a long lived record required not more than some drawer space. In modern mass storage media it is the time the space remains occupied that counts.

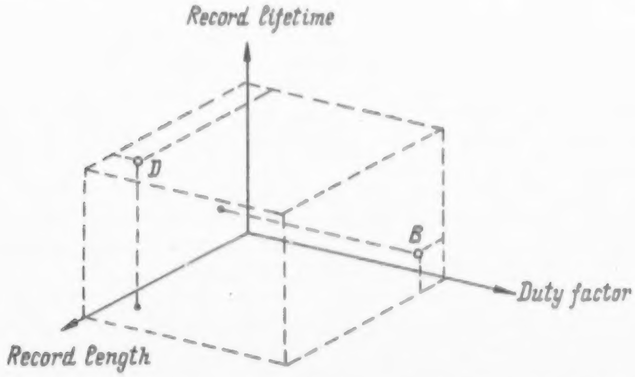
(c) In a business data file a record may vary in length between a few characters and hundreds of characters, whereas the length of a typical documentation record begins where the length range of business data records ends.

(d) As for the item's mean value, in business files the cost of storing and processing each individual record uses to be discounted from the value of the respective item. Needless to say that a business system would scarcely capture an item of a value lower than the storing and processing cost. On the other hand, the «value» of an individual documentation record is very vague and might be determined but indirectly e. g. by averaging in from data derived from the overall effectiveness figure of the file.

When taking business IR and documentation IR Systems as two extreme cases of the same problem, we can illustrate the foregoing parameters in a Table:

Parameter	Business File	Documentation File
Utilization (duty factor)	high	low
Record length	short	long
Record lifetime	short	long
Item value	adequate	?

Any two or three of the above parameters may be represented in a two- or three-dimensional diagram, respectively, e. g.



The respective positions of typical documentation and business data files are labeled D and B. It should be noted that they lie near the opposite corners of a cube formed by the coordinates.

Remark:

A more refined method of evaluating the utilization of an information file would include the average number n_s

of relevant records found within a search. By multiplying n_s by the daily duty factor D and by the expected mean life time t of a record we get what we might call the pageing factor p of the records in the file:

$$p = n_s . D . t \qquad (2a),$$

or

$$p = n_s . q . t / F \qquad (2b)$$

Figures of p met with practical systems show that while documentation files do not score more than about 10^{-2}, a business or data bank file may reach a value of 10^{0} or even more, although the mean life time of a record in the latter is generally short and a typical answer may contain but a single record. In other words, in the business file types it is the high utilization rate of each record which is responsible for the relatively high p values. Each record in a business file is paged many times during its short life while in a documentation file just a minority of information stored is called up a few times during the file's entire long life.

Now let us return to the impact of the above factors on the system's design. Most of our documentation IR systems are sequentially operated when searched, i. e. whole files (or substantial parts of them) are scanned in each search run, independently from how many queries are involved. Such a system necessitates a cumulation of queries, i. e. batch processing.

The problem does not lie so much in the fact that of course real-time processing is inconsistent with cumulating batches, as in the long time needed for gathering a batch big enough to be worth processing at all.

The user feels attracted by up-to-date types of services and appreciates their high speed and reliability as provided by computers. On the other hand, he dislikes giving up facilities he got accustomed to, such as browsing possibilities, immediate or at least short term answering of queries, and parallel searching.

In a typical sequentially scanned documentation IR system the cumulation period may attain as much as several days for reasons of economy; thus the possibilities of man-machine conversation or browsing are severely constrained. Actually, the overall speed of a sequential scanning IR system with query cumulation is lower than the speed of most manual systems of the same size.

Any correction of a query resulting from a not fully satisfactory previous search result — which is a procedure inherent to documentation IR — introduces a delay equal to the cumulation period. Thus the ultimate volume of any sequentially scanned file is determined not only by the technical parameters of the computer, such as magnetic tape length and storage density, transfer rate, etc, and by the retrieval language parameters, but primarily by the duty factor D.

In documentation IR systems — as opposed to business data files — the duty factor figures are inherently low, which is due to the quantitative discrepancy between the needs and the volume of information stored. Several case studies of research and/or development projects have shown that the queries requested are not high in number during the whole project period. Typically but a few large searches were required within several years of research, not counting a lot of complementary or auxiliary searches which, paradoxically, aimed into other fields (e. g. a wiring diagram of a voltage regulator was searched for in a biological research project).

Example:

For a modern medium size computer with fairly fast logic and a quite primitive indexing and retrieval language with no syntactic rules a scanning rate of 10^5 records per hour may be quoted as a representative figure.

Assuming a cumulation period of 24 hours as just adequate and allowing a rather high cost of each query equal to one computer working hour, and neglecting all other variable factors, would lead us to a rough estimate of an economically justifiable duty factor $D = 10^{-5}$. This figure corresponds to about 2500 queries a year for a file of 1 mil. records, which is far more than the figures met with documentation IR systems. On the other hand, it is below the figures representative for business data files.

There is but one way out of this blind alley: some kind of inverted file in a random access memory At the time being, the only applicable storage media are the magnetic disc or drum and the magnetic card. In the near future e. g. photoscopic memories with coherent light read-out might be the answer. For inverted file systems the duty factor is not as important as for sequentially scanned systems because no query cumulation and hence no substantial answering delay is involved. Nevertheless the duty factor provides a rough indicator of the utilization of the system.

With inverted files real-time searching is feasible and no query cumulation is needed. Search time are much shorter than with sequential scanning but the economies in computer time are traded-off for high bit price. Therefore business data files with their short record length, short record lifetime and high item value are more suitable being for stored on RAM devices than documentation files.

Calculating the storage capacity requirements for a file is simple arithmetic. For documentation, at least 300 characters or some 3000 bits per record stored are to be counted with. Thus a medium size file of 100 000 records will require a storage capacity of about 3.10^8 bits which is much but still within the range of RAM stores. What is worse — besides the relatively high bit price of such devices — is the fact, that each bit element in the store will be occupied once for ever.

Therefore, big computer documentation files have to be waited for until the time when the price of accomodating information may be neglecter. This goal is being reached by now by the magnetic tape it is far ahead for other high-capacity and short access-time storage media. The hardware and software do not match the requirements of documentation IR. It dawns to me that the manufacturers of computers have not yet begun to consider the documentation IR market interesting and extensive enough. Otherwise they would have provided this field with more specialized equipment. Up to now documentation retrieval hardware has followed with few exceptions the way tracked by accounting machines.

The above considerations may at least partly explain certain observed facts in the IR field, in particular

- why business IR files are more successful than their documentation counterparts;
- why business IR files are better suited for inverted storage on RAM devices;
- why the documentation services had to look for the way of making indexes as an interim solution which would

have taken the place of the not fully satisfactory centralized IR systems.

In conclusion another inconsistency in the IR practice should be pointed out: While projecting an IR system, the designer tends to make the most of the information that is going to be put into it. In most computer oriented systems, IR retrieval is complemented by some type of current awareness service such as SDI, printed subject and author indexes, and the like, so that each record is disseminated in one or another way before being stored for retrospective retrieval in a more or less distant future. It has been shown that the economy limit of an IR system depends to a high degree on the magnitude of the duty factor, i. e. — for a given file — on the number of queries in a reference time period, e. g., a day. But providing SDI and indexing services — not to speak of reproduced magnetic tapes and punched cards dissemination — is at odds with the needs of a centralized IR service because many potential customers of the IR service are drained off by making searches by themselves in their own files and indexes. The designer of any information system is thus confronted with the dilemma centralized IR or indexing service, while primary economy considerations would press him into a solution IR and indexing service.

This is a point where more detailed investigation into the interaction between IR and its derived or conjugate services is needed. Rough estimates indicate that a half-and-half sharing of the respective duties might be expected. Moreover, the conjugate use of the same format (say abstract) for both types of service is not always efficient because it is clear that an abstract made for current awareness purpose or simply for «news» dissemination cannot be adequate several years later when incidentally paged off a file in an IR search.

R. A. Fairthorne
(United Kingdom)

PROBLEMS OF DATA RETRIEVAL AND DEPENDENT TECHNIQUES

This paper outlines the basic problems underlying Data Retrieval and the hierarchy of techniques that it supports from 'look up' through 'question answering' to 'answer inferring' from records or text. Little of what is said here is novel. The problems described are being studied by many workers. Here I attempt only to sort out the rather tangled issues involved and to use them to put some important techniques into perspective.

A very necessary preliminary is to separate these various techniques from the even more various activities they aim to aid. To omit this is to risk being like those who cannot discuss the building of a bookshelf unless they bring in and discuss the psychology of authorship, the economics and ethics of book publishing, and the social effects of libraries. All these are matters that must be discussed, but not whilst constructing the bookshelf. So here we assume only that, for reasons outside the scope of this paper, it has been decided that techniques should be available for finding segments of what people have had to say, in words or writing or pictures or sounds or other-records, and then to make inferences from these segments. The motives for these inferences do not affect us directly. The modes in which people talk and record and publish for various purposes do affect us directly.

For the same reason we must ignore procedures for verifying what people have had to say and of estimating confidence in it. Although verification and weighting of statements are essential to the proper use of documents, to include them as a part of data retrieval would be circular. For verification and confirmation themselves involve data retrieval and cognate techniques; for example, to find out what others have had to say about the matters retrieved and how reliable they consider the author. Verification by direct experiment or observation of events themselves rather than of what people have had to say about them, lies completely outside the scope of Data Retrieval or, indeed, of Informatics itself. Menus are not the same as the meals they denote, nor are libraries to be confused with laboratories. The function of libraries in particular, and of Informatics in general, is not the pursuit of knowledge as such, but of knowledge of what people have had to say. In short, Informatics is concerned with the management of messages, not with their creation or application.

One key task of message management is to isolate some fragment of a text or diagram or record without compelling the reader to hunt through all of it. Immediate examples are the provision of the numerical value of some function for a specified mathematical function; of times of departure and arrival from and at various places by rail, ship, or air; of a paragraph or sentence mentioning the name of some person, place, substance or event; of the part of a map, or pean or circuit diagram that contains some named part or some specified configuration of conventional signs; of some verse or lines of poetry that correspond to a possibly incomplete or inaccurate recollection.

These are examples of what is called Data Retrieval, an Activity that aims to satisfy the very common requirement for a part of a text or record as opposed to a request for the whole of it. The data retrieved is the actual segment of the text or record, not what the segment may refer to. What one retrieves from a mathematical table, for instance, is not the value of a function, but an expression representing the opinion of the table maker about the value of that function. What one retrieves from an airline schedule is not as is well known, the actual times of arrival and departure but an expression representing the intentions of the air transport organization concerned.

What makes the sequence of digits or the alphanumerical expressions cited become 'data' in the sense of Data Retrieval is the fact that they are required by themselves in isolation from the rest of the text. What makes texts or records 'data' therefore does not depend upon its symbols, wording, or topic, or any intrinsic characteristic, but upon how it is to be used. In linguistic and literary research any document may be used as a collection of data for separate retrieval. Indeed, standard works — Shakespeare, the Bible, major poets — are often arranged and numbered for Data Retrieval; that is, for looking up parts of the document without having to read through the document as a whole.

Data Retrieval, or 'look up' is an essential preliminary to all forms of 'question answering' and other inferential procedures from records, whether automated or not. Data Retrieval is necessary for these, though not sufficient. If you cannot do it, you cannot go on to do the more complex operations. If you can do it, you have still to be able to carry out the more complex operations.

Also, though this is often overlooked, you cannot begin to perform even simple look-up unless you first perform ordinary Document or Record Retrieval. For you must first get hold of the appropriate documents from which the segments of text are to be extracted. But the topic of a document as a whole cannot be deduced from its textual content in isolation, it depends also upon what kind of reader is going to use it for what kind of purpose. Patent Office searches, for instance, try to find prior disclosures of ideas claimed to be novel, wherever and for whatever purposes these ideas may have been described, in fact or fiction. Thus in theory any published document may inadvertently contain a prior disclosure in the light of retrospective search. For instance, in the USA humourous and fictional cartoons have been cited as prior disclosures of patentable ideas subsequently claimed as novel.

The patent offices of all countries have to use the complete armoury of bibliographical weapons to select from world literature as a whole, and to lead down through classes of documents, through single documents, to the sentence paragraph or picture expressing the patentable idea. Clearly there can be no complete solution, even if we were to read every word and look at every diagram and picture ever produced. For we would still be left with the problem of recognizing which of these were relevant. This usually would involve problems of inference from the text or picture in the light of the motives of the searcher and of the original author of the document as a whole.

For the purposes of this paper I will take a 'document' as being any representation of messages that is treated as a unit of discourse in the social environment concerned. Clearly what one person regards as a document may be a part of a document or a collection of different documents to someone else, or to the same person on different occasions. For instance, a librarian will regard the set of volumes comprising an encyclopaedia as being one document, as will an accountant, for if any one of its physical component volumes be missing, the encyclopaedia will be incomplete. An individual reader of the encyclopaedia would regard even one volume to be a physically bound collection of different documents in which the one he wanted to read was embedded. Similarly a set of bound issues of a periodical, a single issue, and one of the articles in an issue may each be regarded as a document according as to how it is used.

Because a document is not a fixed unit of discourse but varies with how the records are to be used, there is no fixed level at which document retrieval ends and data retrieval begins. On the other hand, no single application of a bibliographical method, such as subject classification, can extend unchanged down from the collection as a whole into the ultimate words or signs of the records without changes of criteria. This is manifested by the notorious differences, by no means fully explored, between indexing a collection of books and indexing the contents of a single book. The relations between documents considered as units, and within a document considered as an ensemble of units differ, whatever entity may be chosen to be, for the purpose in hand, a document. At root the problem is that the topic of a document considered as a whole is not necessarily that of the topics of its component parts; e. g. the theme of a novel or poem is not necessarily mentioned in any part of the novel or poem; nor is the topic of a scientific journal the topic of the several articles that compose it. It is only occasionally possible for the topic of a sentence, in the sense of what it mentions, to be the topic of the paragraph in which it appears, or of the paragraph to be that of the chapter in which it occurs. In short, we cannot find out the topic of a document from the topics of its components alone. Even if we had a concordance such as, say, a keyword-in-context (KWIC) 'index' that covered the contents of an entire library as if it were a single huge document, this would be useless other than for purely lexicographic search. A concordance begins its work after the document it serves has been picked out from other documents by bibliographical tools designed to distinguish between documents, not within them. Mathematical tables, which are extreme cases of data yielding documents illustrate this point. Inspection of the entries in a mathematical table yields the numerical values, or finite approximations thereto, of some mathematical function at cited arguments. No inspection, analysis or computation based upon these can determine what the function is. Any number of polynomials or other functions can be found to fit any finite set of finite numerical values for given arguments (Weirstrass' approximation theorem). To find what the function is we must look to the page headings or the title on the spine of the book or other metatextual expressions.

Even more do we have to appeal to book, journal, and article titles, and the intentions of the reader, when assigning topics to fragments of text for the purpose of data retrieval. The shorter a linguistic expression, and the smaller the segment of picture or diagram, the greater the variety of documents of which it could form a part. Therefore items to be used for Data Retrieval must first be classified in some way or other, pertinent to their intended use, before data retrieval can commence. In no way does data retrieval bypass the normal bibliographic procedures. Indeed it demands even more powerful procedures even more powerfully applied.

Documents compiled specially for data retrieval of known purpose by readers of known habits are usually designed to be congenial to these habits. The physical and visual formats attempt to satisfy the anticipated conditions of consultation, so far as these complex demands can be met. On the level of 'human engineering' alone, acceptable satisfaction is achieved only by uneasy compromise between weight, size, typeface, layout, binding methods, and characteristics of the printed surface. Should the compilation be needed for whole or partial machine use, it must be completely redesigned. If it is to be consulted by many people, each for a single retrieval, it must be designed differently than for consultation by few people, each for many retrievals at a time. Again, the design will vary as the use involves a sequence of systematic retrievals, or a set of arbitrary retrievals.

These physical aspects must reflect more than the physical and operational characteristics of the anticipated type of user. They must reflect the structure of discourse within the topic for which the data items are to be used. Layout, headings, titles, and other metatextual devices are manifestations of the logical and semantic structure appropriate to the topic discussed, and the discussants. If the same items are to be used for another purpose, their arrangement and labelling within the document must be altered accordingly.

For instance, single items of data that couple a name and a telephone number will be differently arranged, printed, and indexed for a general telephone directory, a directory of trades and occupations, or for an inverted directory of names corresponding to numbers. The physical characteristics will vary according to whether the directory is to be used by the general public in a public place, by trained staff in an office, or in conjunction with scanning devices. Again, the actual entries in a table of Bessel functions of high order are a selection from those in a table of Bessel functions of lower order or, in some cases, of those in a table of ordinary circular functions. The items are just differently labelled.

Evidently even moderately efficient and convenient data retrieval, even for a sharply defined purpose and clientele, demands detailed design of the data yielding documents and of the ways in which they are to be used. Also the overall subject classifications and other bibliographic tools of the documentary environment must be aimed at data retrieval for the declared purpose. Little evidence supports the comforting belief that an entire collection of documents can be organized in advance on, say, magnetic tape so as to allow direct data retrieval for any purpose in the future. The best we can do is to deploy duplicates of relevant parts of the collection to deal with existing demands for data retrieval, to understand or discover the principles needed for redeployment, and to have the physical and computational resources to do so when needed.

Many documents not habitually used as sources of data can be called upon to do so under the demands of linguistic, stylistic, historical, and political studies. For these the traditional fusion of special format and conceptual structure is not usually feasible. Even in texts, such as the plays of Shakespeare, that are often used both for data—e. g. quotations out of context, stylistic analysis—and for reading as a whole, compromise between format and convenience is uneasy.

All indexing and formatting devices for data retrieval aim eventually at locating and extracting a fragment of text or record, considered at first just as a string of symbols. The conceptual criteria for selecting fragments are built into the index and format. They are arrived at initially from people who have read the complete document from the viewpoint of those who are going to use it as a source of data, e. g. book indexers. Sometimes, as in linguistic analysis, the text strings as such are the objects of interest, and therefore the entire text is to be used as data. Documents that are to be used in this way by enough people may be arranged and indexed with respect to words and their sequences just as are other data compilations. Whether this be done or not, at some stage the entire text must be scanned by some device, human or mechanical, capable of recognizing, locating, and counting expressions of the required types. This task demands much of both humans and machines, even when given rules for identifying the required expressions. On the physical level, recognition of even quite sharply specified patterns, such as printed words in a limited range of size and typeface, segments of musical notation, chemical formulas, conventional signs on maps and circuit diagrams, is by no means theoretically established or operationally feasible in general. Not the least problem is the purely mechanical one of manipulating documents of different sizes, formats, and bindings. Holographic and related techniques based on generalized Fourier synthesis and analysis can pick out patterns wherever they occur on a plane, provided that they are not askew, but such methods are in their infancy. Closely related to these problems of input are those of output, involving synthesis for display or format.

The major problems lie in the formulation of recognition procedures at all levels from pattern to topic, and their manifestation as symbol manipulations. The latter task belongs to computer scientists and engineers. The former does not, and is very formidable. Even should we have a device that could read and identify conventional symbols, we have yet to instruct it to recognize those combinations that represent proper names, substantives, and substantival clauses, amongst others. We need these to find out what a document mentions, a necessary but not sufficient step towards finding what the document is about in the context in which it is to be used.

Without doubt computational linguistics and symbol manipulative devices are helpful tools for data retrieval, and are rapidly becoming indispensible ones. But it must be remembered that they are tools, they are not substitutes.

Data Detrieval for the humanities — pictorial art, music, architecture, history, and literature — differ in no way from Data Retrieval for science and technology, other than in being more difficult. The varieties of representation and of discourse are far wider, and they are not so sharply constrained and disciplined. In response to these challenges, the humanities have made and are making important advances in linguistic (including pictorial and musical) analysis and synthesis, and in their computational realizations. I would urge that Data Retrieval should not be considered as something different for the sciences than for the humanities, but that that both should be considered together as part of the same general activity based on the same general principles. Neglect to do this will seriously delay progress and cause much unnecessary duplication of thought and effort.

Data retrieval in the sense of 'look up' or 'look through' to find explicit expressions, is the necessary though not sufficient preliminary for more complex activities such as 'question answering' and 'answer inferring' from records. Before dealing with these we must first bear in mind that any answer we derive from a document or recorded message is a purported answer. It may not be the correct one. Verification lies outside the scope of informatics. This comment should be obvious, but many contemporary writings suggest that their authors believe, or behave as if

they believed all factual statements to denote facts — even in fiction — or that we or a computer can tell whether a statement is true by looking at it. That this is not so is obvious when applying informatics to the humanities. We have here one more argument against keeping scientific and humanistic applications insulated from each other.
If informatics is considered as applied to one field only, then the principles and concepts belonging to informatics in its own right tend to be infected by alien principles and concepts valid only in the field of application.

Secondly, we must bear in mind that if a document is to answer a question by direct data retrieval, then that document must contain the question as well as its purporred answer. Moreover the appropriate question must be unambiguously associated with the appropriate answer. One cannot infer the question from its answer, because any expression can be the answer to any number of questions. For instance, the expression '1917' is the answer to an indefinite number of chronological questions, and also to an indefinite number of computational queries such as «What is the product of the cube of 71 and the cube of 3?». The answer to any computational query, or to any query involving a numerical reply, is to be found in an extensive enough table of randomly generated digits. Any query about times of arrival and departure of aircraft sometimes may be found in the times of departure and arrival of other kinds of vehicle. The problem is how to recognize them. Sometimes algorithmic procedures are available for recognition, though direct calculation of the answer is usually preferable. For the travel query, and for arbitrary matters in general, it is better to obtain the appropriate timetable by normal bibliographic procedures, and then to look up the data entered against your query, as expressed in the mode laid down by custom or the instructions for use of the timetable. From this point of view the wording and format of documents designed for data retrieval ('reference books') are representations of the questions that can be put to them.

Often such data compilations can be used to answer questions not explicitly catered for by the designers, provided that we have valid procedures for this, and that we do not attempt to manufacture data that is not already implicit in the compilation. A common example is to use a table or schedule inversely to find which argument corresponds to a given value, or which name or title to a given serial number. Whether this is more convenient than to use a ready made inverted compilation of the same data depends upon many things, including the existence and accessibility of an inverted compilation, and how systematic the data may be. In the days of manual desk calaculation that involved looking up of values in many different tables, experienced computors usually found it more convenient and accurate to use a direct table inversely than to switch over to a separate inverted table. They did not take this view when trying to use a telephone directory inversely.

In general any document designed for data retrieval and the answering of a particular set of questions can be used legitimately, if not conveniently, for certain types of questions not catered for originally. A document so used will be called 'answer providing' and the process of using it in this way as 'answer deducing'. Many highly systematic documents provide aids for both direct look up of prerecorded answers and for deduction of answers not explicitly recorded. Mathematical tables usually provide for interpolation between entries, and the amount of instruction and printed aids for interpolation will vary with the resources assumed to be at the user's disposal. Present day tables for human use are different from those published in desk calculator days, and those for automatic machine use very different indeed. Tomorrow's tables and reference documents for humans aided by computer networks must also differ from those of today, and those data compilations that are to be used by the computers as an implicit part of explicit instructions will have to be designed with very careful attention to the type of public they serve. They cannot serve all requirements equally well, any more than can a printed reference book.

Only very systematic matters completely closed under algorithmic procedures are valid topics for answer deduction. By 'closed' is meant that the symbols used not only represent real entities in the field before calculation and symbol manipulation, but also afterwards. Clearly it would be incautious to interpolate an airline schedule between the entries for successive departure times, but it would be reasonable to make rough interpolations between the time of departure as recorded and the time of arrival at its destination.

Though the fields to which answer-deduction is valid are very restricted in scope, they are very important; for example, air traffic control, shipping and cargo deployment, communication satellites and spacecraft, logistics of nursing and hospital management, civic administration of well defined activities, manufacturing processes, military applications and, of course, the logistics and management of libraries. Each such application is highly specialized, and its design demands extremely careful and acute study of the samantics of the situation by subject specialists, of the design and assignment of labels for the topics and structure of discourse by documentation specialists, of the representation and organization of the labelled items by computer specialists, and the physical and operational integrity of the system by engineering specialists. The user also must make some effort to use this tool properly, as he must for any tool from a screwdriver to a computer network or, for that matter, a library.

Though strict deduction and calculation of answers are valid only within very narrowly defined and completely understood applications, these are not the only valid forms of inference. Few conclusions drawn in ordinary life are as strong as those drawn from deduction or calculation. However, weak inferences are by no means useless. It is better to have a weak conclusion correct as far as it goes, than to have a strong but incorrect conclusion, correctly deduced from false or incomplete premises. Most services rendered by libraries and cognate agencies are necessarily based on weak inferences. Not being omniscient and omnipotent, and not being identical with the readers who will use the services, they aim only to provide documents, records, and data that are plausible responses to what their customers ask for. If asked for the answer to an explicit question they seek this by direct data retrieval, or by deduction or calculation from explicitly answer providing records. These may not exist. In this case, the services must seek to supply data, in the sense of fragments of or citations from records, from which it is plausible to assume that and answer can be inferred. This demands knowledge of the background knowledge of the user in this particular case. We must note that this is by no means the same as possesing this background knowledge itself, as possessed by a practitioner within the topic concerned. Nor is it, nor can it be, the knowledge possessed by the individual making the request. What it is 'book knowledge'; that, is, familiarity with the modes of discourse enough to recognize passages of text from which the expert reader might infer the answer he seeks. There is no certainty that he can do so, or that he will do so. The procedures aim to make it reasonable to assume that he may do so.

It is well known that this limited aim is not only often achieved, but also is well worth achieving. A trained librarian, or services designed by him, who can indicate documents or data that probably will interest or inform particular readers does more than just save them time and trouble. He may trigger their imagimation and curiosity. The problem here is how to delegate such 'inferential data retrieval' or 'answer inference' to less trained agents or to automate wholly or in part, or to make it possible for a trained librarian to do more of it for more people in a given time with given effort.

In special case it may be good enough to search passages of text or parts of pictorial records for the presence of particular names or symbols. To make such search worthwhile, it must be more than good enough, unless very weighty matters are involved, or the users are prepared to submit the output to very refined examination or the topic and mode of discourse are so narrow that readers' interests are adequately indicated by the occurrence of particular words or symbols. Searching of records, by humans or machines, is a difficult and expensive operation.

Even if complete searches of text are feasible, search and selection for inferential data would still demand knowledge of the structure and language of discourse of the potential reader, of the author of the text, and of how to correlate the two. But these are also necessary for the efficient design and application of documents made for explicit data retrieval. Also, of course, they essential for subject classification, 'content analysis' and other operations aiming at document retrieval as opposed to data retrieval.

The two are not in fact opposed, one deals with documents as a whole, the other as with fragments of documents. But we have seen earlier that the notion of 'document' is relative to use, not an absolute property of an isolated block of recorded discourse. The differences between document and data retrieval arise only from different choices of unit of discourse or record. From the point of view of question answering in the wide sense, they are both tools of Reference Librarianship in the wide sense. They must not be regarded as distinct, but as governed by the same principles, and as liable to merge one into the other, whether the records used are mathematical tables, musical scores, maps, sound recordings, or novels.

From the physical and engineering points of view the differences between dealing with an entire library or with a fragment of text are considerable. Transcriptions and formats must be designed for particular needs, they cannot satisfy all of them at once. Automation for data retrieval initially should be designed as the personal tool of a trained person, corresponding to the traditional Reference Librarian, acting as a delegate of the readers. Automation of data retrieval and its derivatives for the direct use of readers, as in a multiplex computer network, entails automation of the reference librarian as well.

The root problems of Data Retrieval remain those of Document Retrieval, both in the bibliographic (descriptive) and semantic (topic, subject) aspects. We have to find how much social and linguistic experience of various activities must be built into the system, and also to find how to build it in. These problems are common to applications in both the sciences and in the humanities.

R. Leclercq and **C. Rixon**
(Belgium) (United Kingdom)

A CONDENSED HEURISTIC DOCUMENTATION SYSTEM

1. Introduction

The scientific world is currently alarmed by the existing exponential rate of increase in documentation. Consequently, a system which would allow condensation of numerous scientific data, literature and experimental knowledge in one complete treatment is of undeniable interest. If such a system also holds the possibility of suggestion of ideas and even of organisation of research, then one obtains a process which constitutes a serious step towards a solution of this major problem confronting scientists. It is this process which we present under the name of Condensed Heuristic Documentation (C. H. D.)

Our system has perhaps one drawback in that it assumes a good analysis of publications and a precision of reasoning. This inevitable obstacle of the actual form of presentation of scientific texts could be overcome in the future if scientists were to agree on a more consistent use of words in a more precise language. Science must find its language and, until it does, researchers will have to rely on specialists in documentation to perform a part of their work.

The system proposed is based partly on a generalised logic described by one of the authors [1] and partly on a consideration of couples of concepts and the transition from couples to series, two ideas which are fundamental to the human mind and which have been fully described by the psychologists Henri Wallon and Jean Piaget [2]. The idea of couples is also convenient since it allows a simple representation in 2 dimensions and the use of classic perforated card indices.

2. Coupling of concepts

Our method is independant of the choice of concepts which must only obey the rules of their own particular science.

Concepts are considered first of all by couples, each couple of concepts being connected by one or more **coupling numbers** which express in a conventional way, all or part of the following indices:

reltaion existing between the concepts considered,
plausibility of the relation,
weight of the 2 preceding indices, and bibliographical **reference**

Each of these terms will be defined in paragraph 3.

The coupling number can be written in any system, whether decimal, dual or other, the transition from one system to another being classic. A number described in the decimal system can therefore be converted to a number of base 2 with a view to its use in a computer programme.

A possible example of a coupling number in the decimal system is:

5412	/ 31	/ 07	/ 45
Number of the bibliographical reference	weight	plausibility	relation

Such an expression is valid only if the same number of figures always refers to the same classification of terms and if codes for relation and plausibility are established.

In the example given, the 2 figures on the right express the relation and the next 2 figures to their left, the plausibility.

If one is not interested in the weight but can limit oneself to plausibility, the 2 corresponding figures can be omitted, thus simplifying the expression. Similarly, if one does not wish to retain the references, one omits the figures on the extreme left.

3. Definition and calcultaion of indices

a) **Relation** is the nature of the logical link existing between two concepts. There are several types of relation, of which it is customary to consider first **inclusion.** This is represented in most classic logics by the symbol $\subset$, where the expression $A \subset B$ corresponds to $\left[\boxed{\oslash A}\right]^{B}$

If $A \subset B$ and $B \subset A$ we can write $A = B$, giving us the relation of **equality**.

Implication is usually represented by the symbol $\rightarrow$ where $A \rightarrow B$ means that A implies or leads to B. If $A \rightarrow B$ and $B \rightarrow A$, then we can say that A is **equivalent** to B.

Further, if $A \rightarrow B$ and $C \rightarrow B$ then A is **analogous** to C. One can, of course, imagine other relations and conditions as well as those defined, a further example being the relation in experimental science expressed by the words «has a value of», which allows the coupling of a general concept with its measured value.

b) **Plausibility** is the value attributed to a judgement. As shown elsewhere (4), plausibility can not be considered as a probability, since it does not satisfy the condition $pA + p\bar{A} = 1$.

The following discontinuous scale of plausibility is employed:

b)

—1	False
—0.9	very probably false
—0.7	probably false
—0.4	hypothetically false
—0.2	very hypothetically false
0	uncertain
+0.2	Very hypothetically certain
+0.4	hypothetically certain
+0.7	probably certain
+0.9	very probably certain
+1	certain

Plausibility can be determined according to the following rules which may take the form of a definition, or a convention translating as closely as possible a line of thought, or a compromise with the theory of probability, and which, all together, must lead to a consistent theory:

1. If the same relation between two concepts is expressed bu different plausibilities and weights, one can condense the information in a single coupling number which has a plausibility value of $\frac{\Sigma P_1 . P_i}{\Sigma_i}$ where P_i is the weight of the various individual coupling numbers and P_1 the corresponding plausi-

bility. In this case, we have a definition of plausibility which we treat as a means of measurement.

2. If A is analogous to B, and B is certain or very probably certain, then we can say that A is very hypothetically certain. Here the rule takes the form of a convention.
3. If A is analogous to B and to C and both B and C are certain or very probably certain, then A is hypothetically certain.
4. Rules 2 and 3 are valid if the word «certain» is replaced in each case by «false».
5. If $A \underset{\subseteqq}{\rightarrow} B$ and $B \overset{P'}{\underset{\subseteqq}{\rightarrow}} C$, then $A \overset{P.\,P'}{\underset{\subseteqq}{\rightarrow}} C$ i. e. the plausibility of the final relation is the product of the individual plausibilities of the two initial relations. This is a convention based on the transitivity of the relations — implications, inclusion and equality.
6. If $A \overset{P}{\rightarrow} B$ and $A \subset C$, then $C \overset{P'}{\rightarrow} B$ where P′ will be very hypothetical for any level of plausibility, p. This is a convention of generalisation.
7. Concepts which are equal or equivalent have the same level of plausibility.
8. 'Students' rule of t' (3) is compatible with our discontinuous scale and our definition of plausibility for $t<1.5$, as has already been established in a previous work by one of the authors (4). This rule involves a compromise with the probability theory, and can be used to calculate the plausibility of quantitative observations.
9. The plausibility of a qualitative observation must be assessed by the scientist in accordance with the usual criteria and will normally be «probably certain» or «very probably certain».

There are other minor rules governing the determination of plausibility (4) but those given here are the most important.

c) The **weight** of a judgement expressed on a relation between two concepts is the value of this judgement with respect to all previous judgements on the same relation between these two concepts. In practice the same weight (i. e. 1) is often given to each judgement of the same relation, and if the plausibility figure which appears in the final coupling number results from, say, 10 judgements, the weight of this plausibility, p_1, will be 10. A single new judgement of plausibility, p_2, will have a weight of 1, but on adding it to the system, we obtain a final weight of 11 and a corresponding plausibility of $\frac{10 \times P_1 + 1 \times P_2}{10 + 1}$ (rule 1 § 3b)

d) The **reference** is that used classically in documentation for easy retrieval of the document which has been used as the basis for writing the coupling number. It is often the actual number of the document in the chosen system.

4. Transition from couples to series

The transition from couples to a well-ordered series of concepts must first satisfy the rule of transitivity:

$$A \subset B \text{ and } B \subset C \text{ leads to } A \subset B \subset C$$

It must then satisfy the rule of coherence:

$$\text{If } A \subset B \text{ we can say } B \rightarrow A$$

The seriation of concepts from couples connected by the same transitive relation is a perfectly mechanisable operation. This can be illustrated by the following easily generalisable example:

Starting from

$$A \rightarrow B \quad (1)$$
$$B \rightarrow C \quad (2)$$
$$A \rightarrow D \quad (3)$$
$$D \rightarrow B \quad (4)$$

we can write

$$\text{from (1) } A \ldots \rightarrow B_1 \ldots \quad (5)$$
$$\text{from (2) } B_1 \ldots \rightarrow C_{11}. \quad (6)$$
$$\text{from (3) } A \ldots \rightarrow D_1 \ldots \quad (7)$$
$$\text{from (4) } D_1 \ldots \rightarrow B_{11}. \quad (8)$$

and from (6) we can deduce:

$$B_{11}. \rightarrow C_{111} \quad (9)$$

and hence finally

$$A\ldots \to D_{1}\ldots \to B_{11}. \to C_{111}$$

We could just as easily have written:

$$A\ldots \to B._{\cdot 1} \quad (5)$$
$$B\ldots \to C_{\cdot 11} \quad (6)$$
$$A\ldots \to D._{\cdot 1} \quad (7)$$
$$D._{\cdot 1} \to B_{\cdot 11} \quad (8)$$

and $$B_{\cdot 11} \to C_{111} \quad (9)$$

which arrives at the same result.
We have very simply allocated indices to the concepts, albeit as basically as possible, but in such a way that the order existing between any two concepts is indicated by the indices.
Where concepts are connected by different relations, or by equivalent but intransitive relations, the problem becomes more complicated. It is often preferable in this case to replace seriation by a matrical representation, which can then be easily converted to a linear form. A simple example of this process is the following:
the following:

	A	C	D
B	⊂	→	
C			→

corresponds to $B \subset A$
and $B \to C \to D$

from which we can deduce: $A \xrightarrow{P'} C \to D$ (rule 6 in § 3a)

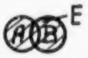
5. Operations on concepfs

The usual operations to be considered on the concepts are logical and mathematical. The mathematical operations are well known and will not be discussed here. The logical operations are union, intersection and abstraction and satisfy the following definitions:

Union of A^{P} and $B^{P'} = A^{P} \cup B^{P'}$

$= E^{\frac{P+P'}{2}}$ (for A and B having the same weight—rule 1 § 3b)

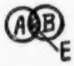
Where E is the union of the intersections and may be represented pictorially as follows:



Intersection of A^{P} and $A^{P'} = A^{P} \cap B^{P'}$

$\sim E^{P'}$ (for $P' \leqslant P$)

E represents the intersection of the extensions



Abstraction of A^{P} and $A^{P'} = A^{P} \& B^{P'}$

$= E^{P}$ (for $P' \leqslant P$)

E = intersection of the comprehensions and union of the extensions.

It should be noted that the extension of a concept refers to the number of objects which the concept signifies and the comprehension refers to the qualities of the concept. Using the system described for the representation of concepts by couples, new couples can be formed starting from a series of couples, for example:

$$\left.\begin{matrix} A \subset C \\ B \subset C \end{matrix}\right\} \begin{matrix} \text{gives us } (A \cup B) \subset C \\ \text{and } (A \cap B) \subset C \end{matrix}$$

New concepts created by combinations of initial concepts are designated by a word or figure*. The retrieval of these combinations may be simplified by constituting perforated card indices (or auxiliary memories) which give a list of the various combinations carried out on the initial concepts.

In certain cases, it is necessary to establish a list of categories for the combinations made. It is difficult to imagine, for example, a complete list of carbon compounds. Here, one would have to consider first the main classes of products or some other criterion.

Conversely, an initial concept can be resolved into its elemental concepts which may be studied in the same-way as combinations. In this case, the elemental concepts or combinations of these elemental concepts can be deduced from an analysis of the initial concepts.

---

* or a combination of letters and figures.

6. Calcultations on coupling numbers

We have shown how coupling numbers can be calculated for one couple of concepts. Several coupling numbers of similar relation can be replaced by a single number having the same relation but a plausibility calculated by the formula already given: $\frac{\Sigma P_i P_i}{\Sigma P_i}$. Different relations for the same couple can be combined according to certain rules.

As we have seen $A \subset B$ and $B \subset A$ leads to $A = B$,
$A \rightarrow B$ and $B \rightarrow A$ tells us A is equivalent to B,
$A \rightarrow B$ and $C \rightarrow B$ tells us A and C are analogous,
$A \rightarrow B$ and $A \subset C$ gives $C \xrightarrow{P'} B$,
and $A \underset{\subseteq}{\rightarrow} B \underset{\subseteq}{\rightarrow} C$ gives $A \underset{\subseteq}{\rightarrow} C$.

Combinations are logical or mathematical operations.

7. Scientific reasoning

We have already seen some aspects of scientific reasoning in analogy and generalisation. Two others which deserve a special mention are the formulation of hypothese and the «game theory».

A hypothesis results from a combination of factors which are each obtained by analogy, the combination itself also being an analogy. A simple example is the case of an unknown product, X, giving the decomposition products, A, B, and C:

$$X \rightarrow A$$
$$X \rightarrow B$$
$$X \rightarrow C$$

If we know the factors, M and N, which give the same decomposition products, we can deduce that X is analogous to M and to N. Further, if M and N have in common (by abstraction) a structure, S, we can say that X has a plausibility, p, of having the structure, S. This can be represented in matrical form in the following way:

	A	B	C	M	N	(class of structures, S)
X	→	→	→	analogous	analogous	⊂P
M	→	→	→			
N	→	→	→			⊂
MaN						

The game theory has been applied in scientific research for the organisation of experiments. Its matrical representation is classic, being that of the hypotheses, H_1, H_2 and H_3 and the corresponding optimal experimental plans, pl_1, pl_2 and pl_3, which may be the result of an analogy or a statistical rule. This is written in part by the general matrix:

	a	b	c
$H_1 \cap pl_1$	→		
$H_2 \cap pl_2$		→	
$H_3 \cap pl_3$			→
$H_1 \cap pl_2$	→		

where a, b, c, etc., are the results of application of the various combinations of hypotheses H_1, H_2 and H_3, and plans, pl_1, pl_2 and lp_3. The choice of the best plan is made according to the usual rules of the game theory.

8. Example

We have chosen a simple example of application in a field which is familiar to us. The coupling numbers are formed as follows:

1. The 2 figures on the extreme right express the relation according to the following conventions:

01 = →

02 = ←

03 = equivalence

04 = $\subset$
05 = $\supset$
06 = equality
07 = abstraction
08 = «has a value of»
09 = analogous
and so on

2. The third and fourth figures from the right express the plausibility according to the following conventions:

01 = − 1
02 = − 0.9
03 = − 0.7
04 = − 0.4
05 = − 0.2
06 = 0
07 = + 0.2
08 = + 0.4
09 = + 0.7
10 = + 0.9
11 = + 1

3. The fifth and sixth figures from the right express the weight which will always be less than 99.
4. The reference, on the extreme left, is the number of the document consulted or the number given to the observation or to the experiment if these are classed in the same way as documents.

The following matrix is obtained:

	SO_2	Se	SeO_2
S		101.10.09.09	
$S \cap \text{calories} \cap O_2$	102.02.09.01		
$Se \cap \text{calories} \cap O_2$			103.01.07.01 104.04.09.01 105.10.09.01
			106.15.09.01

where reference number 101 = number of the document consulted
102 = number of the experiment giving this result
103 = the reasoning used
104 = the experiment
105 = further, very careful experiments
106 = calculus of coupling numbers

Mathematical examples, which at first sight may seem particularly difficult, can also be treated quite simply by this method, in which case the plausibilities are generally limited to +1 and —1 and the weights generally omitted.

9. Conclusion

The process of Condensed Heuristic Documentation enables information to be recorded in the form of couples of concepts linked by logical bonds whose value is expressed by a coupling number. It allows the transition from these couples to series or tables which condense the information. This information can then be further condensed by calculations on the couplng numbers. The process also allows reasoing by analogy, abstraction and generalisation at the same time as deductive reasoning.

The original concepts may be easily traced. One way is to use a perforated card index for each couple and an auxiliary memory to give combinations.

REFERENCES

1. R. Leclercq, «The Use of Generalised Logic in Technological Forecasting» — 'Technological Forecasting' 11, 2
2. H. Wallon, «Les Origines de la Pensée chez l'Enfant». Paris, P. U. F., 1945
J. Piaget and B. Inhelder, «La Genèse des Structures Logiques Elementaires» Delachaux et Niestlé — Neufchatel, 1959
3. H. Jeffreys, «Theory of Probability», Oxford, 1939
4. R. Leclercq, «A. Mathematical Logic of Discovery» (Under print)

J. E. J. Halkin
(Belgium)

PROPOSAL AND WISHES FOR AN OPEN STRUCTURE IN THE COMMUNICATION OF INFORMATION

Theme of this work

One can consider with fear the present and increasing predominance towards the centralization of information into more and more powerful knots. This situation may imperil the development of specialized information centers which, solely, are efficient for maintaining the necessary dialogue with responsibles of the often very narrow fields of human activity. It could become painful to keep the good balance between present tendencies, that is between centralized concentration and specialization.

By further improvements of communication we may ease the setting up of networks and interconnections. If these networks possess a free flowing communication, this will implicitely ease a decongestion of encumbered centralized knots and of course improve their efficiency.

Centralized information presents an analogy with centralized management of a company and, on the other hand, the network evokes ideas of dialogue and assignment of authority. This calls for a fundamental change in the structure. The analogy also exists in the results of such a change: the new structure leads to decongestion.

But the communication inside the network can only become a free flowing one if the needed ADDRESSES are unequivocally worded and standardized. These addresses are examined with some detail.

Information parts in two kinds of functions, i. e. the «treatment» and the «communication». However these two functions are very often under a one-headed responsibility.

The setting up of the network needs a centralized responsibility of the communication and independance from the responsabilities of treatment. But on the other hand, the last ones will earn to be decentralized.

A high level of the quality of communication can transform information activity into a trade or a business, a prodigious dynamic and prosperous one indeed.

> «How to imagine answers which solve and simplify? These are problems to which a reflection upon congestion may help».
>
> 1. p. 16]

1. Looking backward: the intellectuals private library

Lets return in time to the period when each intellectual patiently created his own library, as a real external memory connected «on line» to his brain, the only computer then available.

The efficiency of such a private library rested upon particularities which eased communication, particularities even more significant as we no longer succeed in using these same advantages. Here are the characteristics:

— ADDRESS OF THE USER is solely: the proprietor's own.

—ADDRESS of the DEPOSITORY: is also one only, these being a concentration of books in one single place.

— These books were selected with care over years and are well known to the proprietor. Memorised knowledge of BIBLIOGRAPHIC ADDRESSES is good, even if use of an index is a pretty good thing.

— Content of these books has been mastered thoroughly. The photographic memory helps, marginal notes, remarks and underlining with pencil are warning devices. The brain of the creator of this library disposes of fulgurating means of access to information which concerns the INTELLECTUAL ADDRESS specified by the terms (actually we say keywords or concepts) of the given question.

— ADDRESS of the SPECIALIST well versed, whose judgment serves in orienting readings and researches, also is the proprietor's one.

Such a library is a device enjoying free flowing communication, assured by well mastered ADDRESSES. Their interjoinings seem so natural, reflex and implicit that one needs a careful analysis to persuade itself that these five types of addresses have each one their role in reaching the result: a well oiled operation.

A modern information centre established for the treatment of large collections seldom succeeds in attaining such a privileged position.

2. Babelism of systems

Congestion jeopardizes today the former and comfortable situation. Every man in charge of a responsability at any place in the information chain knows that his memory and his judgment are tools requested for helping the work of any «information system». And he also knows that such human components become less efficient when the size of collections increases. Thaz man surrenders his role to systems, and everything rests upon the conception of these.

Recently, in April 1969, editors of the French publication «Prospective», Paris, published a booklet devoted to «L'Homme encombré», the encumbered man [1]. One chapter by Pierre Massé gives three remedies for the present congestion of human life: élimination, ogranization and invention. The present paper makes proposals that may enhance organization.)

Which are the causes of information congestion? The main one is of course the terrific growing of the bulk of published works. But an unexpected obstacle came from the autonomy of each link by which we have specialized inside the information chain, for instance: publishers, booksellers, librarians, responsibles of any information centre and also consulting engineers in any fields. Each of these activities are free from the others: the man in charge of one is not paid to organize free flowing communication between them, he does not care for that as a chief responsibility. Each one will take care of communication inside his own house, but will invite outsiders to adapt themself to his own communication state. The author acknowledges that, likewise other people in charge of a centre, he was obliged to organize his own autonomous ways of communication, but he still deplores such a «babelism» of private systems.

From that moment and as a result of the «babelism» of systems, the growth crisis becomes a communication crisis: general communication becomes intricate, slows down, may even stop.

New necessities are facing the inadequacy of present structures. An information centre which could be competent in any field is an obsolete pattern. The new pattern we need involves to help one another inside a network of centres where the most qualified one has the privilege to answer. All that supposes a free flowing communication which the autonomy of systems does not and can not allow. That situation gives birth to a conflict between old and new patterns. One have to frame a new structure.

3. Obsoleteness of the present structures

One of the main duties for anybody in charge of an information centre is to choose, work out and uphold for his own use the most nicely fitted system of treatment. Here «treatment» means also «communication», and in such a way these systems radiate as feudalism: each system possesses his fief, inside which communication may be a good one, but between the fiefs, that means between systems, communication earns what earns compatibility of systems, often not very much or nothing at all. Constraint to «translate» one system into another is prevailing, need of a «language» understandable by everybody in the chain, more, a language convenient for the computer, become an obvious request.

But the each day reality may be worst. Non compatibility of systems kills mutual acquaintance. Each document has to be registered in each system, bringing evident redundancy of efforts. In spite of a natural willingness mutual help is inefficient. Duplication of treatments is a heavy loss of efficiency, but nevertheless it is easier and less costly than to translate from one system into another.

An information centre may no more pretend to be omniscient or omnicognisant. To help one another thanks to a free flowing communication between complementary activities became the requisite for achievement. That involves compatibility of systems and setting up of communication rules.

To work out and uphold such a general communication urges that a prevailing authority becomes responsible for it, which means an authority above feudality of systems. The autonomy of anybody in charge of a centre, that is his right to choose his own system, requests an accomodation. It is not easy to replace loneliness of the centre inside this fief by an open «one-another-help». That needs a change of structure which may be understood as some loss of autonomy. Nevertheless any responsible of one «link» inside the information chain ought to accept such a reduction of freedom, but, on the opposite side and thanks to standardized ADDRESSES, efficiency of communication would be enhanced. From that moment automatic (and already living) equipments would take up a prodigious mass of tedious, small and tiresome jobs which are still minutely done by men with unnumerable amounts of redundancy in the world. So we see now that: **babelism of systems becomes the destroyer of systems.** Upon such terms we may propose a better understanding of all the information activities. These are everywhere parted in two distinct functions, one of TREATMENT, and the other of COMMUNICATION. The first is the «message», the second is the «transfer», but we speak so easily by using words in metonymy that we confuse the message with its communication. Nevertheless such parting appears quite often in the reality of our structures and of our responsabilities. Wireless or phone is a process for transfer of messages. Bank account numbers are addresses for communication, but responsability of transfer and responsability of management are different: the first belongs to the banker the sacond to the titular of the account. The road traffic code concerns communication, but neither the vehicle nor the transported wares. The use of postal sectors numbers makes our mailings quicker and safer, but that does not modify the «message».

We may lenghten such a list at random, the list of human activities parted between one responsability of TREATMENT and one of COMMUNICATION, and having done so with success and for better efficiency. It is quite amazing that, at the present age of the computer, information transfer is probably alone not to have chosen such a twin-headed structure.

COMMUNICATION appears everywhere with the character of a collective or centralized responsability or authority and, if that is of general acknowledgment, TREATMENT may be decentralized without any risk to become inefficient. Such structure is a combination of a central and collective organization of COMMUNICATION with a functional de-centralization of TREATMENT in an open network.

Urgency of such a structure is born as a consequence of growth of information activities: we needed to diversity and we have ramified in the numerous «links» forming the «chain» of information, but with an effect to slow down communication. That evolution, complexifying with functional diversification is similar with the evolution of management within a company. In both cases the necessity of the dialogue appears, the necessity of a better «one-another-help», and the one of assignment of authority and responsability. That needs, better communication, networks and informal interconnexions. «Between the private man and the community, relations will become more and more numerous, and we need a cybernetic mechanism for a better fitting up»*). These cybernetic mechanisms will use codes, as the road trafic code does at the level of the community. Information transfer may not escape such a natural law, a law thanks to which a new structure of communication is needed when treatment complexifies. Till now transfer of information does not care of that. But such an acceptance is the condition for a more free and more efficient TREATMENT of information, a treatment then becoming able to transform itself in a prodigious industry and trade.

Efficiency of such an organization for a free flowing communication requests compatibility and coherence of the ADDRESSES.

It is not easy to make out such addresses for, indeed, the target is **to promote a general use of them.** The basic terms for such a drawing out will be:

---

* Translated from [1, p. 19]: «Entre l'individu et la collectivité les relations seront de plus en plus nombreuses, et il est nécessaire qu'un véritable mécanisme cybernétique vienne assurer la bonne articulation».

— To draw the functions to be filled by these addresses.

— To create procedures which may ease the use of them.

— To choose each time arrangements convenient not only for the human user but also for usage by equipments without requiring a translation.

— To try to make the addresses mnemotechnical.

— To do so that the addresses may be used for mailing when useful.

A timetable of approach for that problem may be:

— To review what is already done, or presently under project.

— To describe the different jobs to be made.

— To give a priority order for the different practical stages.

— To define the responsibles of the communication's code project and maintenance, and to define the authority to head that organization.

— To realize the project in the frame of a pattern inside which speed and efficiency of communication prevail.

In Chapter (1) we looked at communication inside the intellectuals private library. That led us to the definition of FIVE addresses having each a functional role. Our capacity to organize an international network working with a free flowing communication depends on — or, is similar to — our capacity to enlarge the use of these FIVE addresses up to the international level.

Our target is to transpose at international level the terms that make efficient the use of a private library by his owner.

4. The addresses which requested at international level

We will analyse each type of address one after the other. In each case we will try to discover the procedure going to the most extensive coordination and to the most easy-to-use standardization. But standardization is not compulsion. We may gain a better acceptance if these addresses respond to natural and simple rules, quite unequivocal rules. If we do not proceed that way the standardization may become unpractical, going easily to failure.

The most significant addresses are the INTELLECTUAL and the BIBLIOGRAPHICAL ones. They typify respectively the responsibility of the information scientist and of the librarian. We will analyse both before looking at the others, which are more similar to simple mailing addresses.

Intellectual address

This intellectual address — or intellectual destination — of each document is usually set up by a list of «concepts» which are controlled by a thesaurus. Those thesauri include generally an enormous part of quite specific words of which the use and the meaning are out of reach for most people. They are words for specialists of narrow fields. Stability of such words may be short in duration and also in precise meaning. Such practice to use specific words arises from our academic time and is going on further with our professional and specialist activities. We lose, if never we get it, our taste for synthesis. Our above-all analytic education is a real handicap for the treatment of information. We incline to omit use of generic concepts which may specify the character of a document. Our mind does not organize easily the natural linkage between generic and specific words.

Our first remark is related to the use of specific vocabularies. They ought to be left for use by acknowledged people in touch with each specific field. In view of an information retrieval target it is most often useless to define with the most specific concepts what may be called for with trivial words by non-specialized people. We realize, and thereof the information scientist is very conscious, that documents are to be «indexed» with concepts chosen at several levels, which means concepts covering the meaning of words fanlike from generic to specific. Specialized information centres take place at the specific end of that fan, and centres which receive an encyclopaedical or omniscient mission reach the generic end. The intellectual communication in such a large network of information centres belongs to special linkages that have to be realized by an unequivocal generic-specific structure of thesauri. A deep understanding of these relationships is basic for information communication.

A second remark concerns communication between specialized fields of different kinds. When a specialist escapes in such and other field of knowleage he uses simple words, not the esoteric words of that field, but words of generic feature used by a man of extended general education. He wishes to receive answers to his questions with words of the same feature. The difficulty to communicate between specialists of different fields may become terrific because of the specificity of words. We suffer a growing need of more didactical papers speaking with generic and simpler words.

Communication between different specialized fields may not be easily set up by specific words, but more easily by words of generic appearance which create links at a higher level of ideas. We can conclude that the making of a so called «universal thesaurus» is an enterprise without any hope, and may be also a useless one, for nobody will ever use it up to the very specialized words it will contain.

On the other hand, a single thesaurus controlled at the level of generic words would be a fundamental objective of a different kind, and would be indeed an objective more easy to reach. Such a thesaurus would retain two qualities of exception, to be short and to be stable. That thesaurus responds to the idea of synthesis of knowledge, while specialized thesauri have no capacity to emerge above our current analytical behaviour.

The American committee SATCOM [2] tried to organize information needs in groups supplied by special services, the so called «NEED GROUP SERVICES». The SATCOM studied the U. S. A. case and concluded to the existence of about a thousand of those groups, which is not a very high figure. It seems reasonable to think that a worldwide census will not give a very higher one. The most useful and practical proposal of the SATCOM report was to organize feeding of each group by a kind of surveying made by special information centers. That idea is sane and we ought to make the most of it. We may ramify several levels downward, what means introducig some logical hierarchy in the «need groups» list. That last option may be compared with the head-categories of the «Engineer-Index», or of the «Chemical Abstracts», or with the titles of the secondary reviews of the VINITI, Moscow. Still better, it seems that the. EXCERPTA MEDICA FOUNDATION, Amsterdam [3], took more advantage of that same idea with a four steps logical classification. The last starts from a set of 33 main categories ramified deeper in a logical hierarchy. Presently the whole includes about 3.000 categories and subcategories. Those categories respond to a pragmatic target: to define user's profiles for each document. The indexing of a document in one or several categories grants it a generic definition, which is not always very easy to do with most of the specialized thesauri now at disposal. After that granting of categories, the indexing is developed deeper with use of a very large thesaurus.

The following idea is emerging from the above considerations. For retrieval's sake the intellectual definition of a document requires at least two levels:

— Belonging to one or several categories.

— A coordinated indexing which uses concepts of a thesaurus.

We may agree one with the other, and also at worldwide level, on the definition of these categories. On the other hand the most part of the concepts will positively remain under the jurisdiction of specialists: it is out of sanity to control specific concepts at a general level. Such a position seems to be the one of EXCERPTA MEDICA FOUNDATION, as indeed the granting of categories is made by generalist physicians while indexing by concepts is finally under control of the specialist physician who makes the abstract.

Such partition of the INTELLECTUAL ADDRESS — or, intellectual destination — in several steps is of course a sane practice. Indeed the «categories» represent what is more permanent in meaning and in duration, what is also assimilable by the community, and most of all, what is necessary to address the documents to acknowledged specialists, while — on the other hand — the «concepts» may require all the science of the specialists to be explained in their meaning and in their evolution.

The way is open naturally by such partition for an agreement up to international level. One has first to define the «categories» which represent each «need group», next one has to ramify these categories in the respect of the «profile» of each need group. Such procedure will introduce a lot of more or less generic «concepts», for they are used by several «need groups». But one must keep clearly in mind that most specific concepts will never find a place in the generic thesaurus: they are subjects for specialized thesauri. There is a border closed for specific concepts.

The information scientists have from all times made a partition of their activities in more or less arbitrary categories, each of them being an image of a group of users or of a particular type of questions. What is new in the proposal of this study is to draw such a partition into categories that will become of general accepted use.

The procedure which is here proposed consists in starting at generic level, and then in going down with care to more specific ones. The usual procedure goes just countercurrent; one has built numerous specialized thesauri apart one from the other, sometimes even indeed with a real success, but one failed obviously to realize an integration of such dissimilar material to more generic levels. It is the merit of organizations like SATCOM, EXCERPTA MEDICA and probably a few others to initiate the indexing by the generic end.

To use «categories» at the generic end and «concepts» by coordinated indexing at the specific end is the marriage of two procedures which contend for supremacy since years. Each procedure possesses its trumps. It is more efficient to use these trumps than to persist in a rivalry attitude. Such a combination of both procedures will be a success if we keep clearly in mind the fundamental difference between a «generic category» and a «specific concept». A «category» represents a human activity, a «need group». By contrast one may say that «concepts» are nearly facets inside the need groups.

One must say that the proposed categories have nothing in common with a decimalization or with an arbitrary or compelling hierarchy as the U. D. C. for instance does.

Bibliographical address

Already a few years ago some people — and among them the author [4] — attempted to promote the use of the «CODEN» code of the A. S. T. M. [5] in order to give a bibliographical address by care and under the responsibility of the editor. That address would be unequivocal and quite easy to use by computers. The CID/Euratom [6] tried to stimulate the use of such an address by editors of the nuclear field. For instance the «Euratom Bulletin» still uses the coden «EUBU» to define a bibliographical address from the moment of publication, the same with «ATKE» for the German periodical «Atomkernenergie». These attempts, of which the usefulness was however proved, failed almost completely at that time. However the benefit of such an address for the man and for the computer is obvious. Nevertheless the CODEN was used to draw up addresses for the abstracts published by the secondary literature, as example by the Biological Abstracts, but one never again proposed that editors themselves would give such an address under their responsibility from the publication. That would be the sole procedure in order to standardize the computer input.

The idea of matriculation by care of editors followed another course. American books are matriculated by year in a continuous file and that number is printed by the editor: that is the «Library of Congress Number» of each book. In Great Britain, Professor Foster, Dublin, has realized a complete system for matriculation of book by the editor. That project is now well known as the ISBN (International Standard Book Number). Presently the IFLA (International Federation of Library Associations) works on a similar project.

But the objective is not matriculation of articles published in the periodicals or of books. What the information community needs is a bibliographical address looking like a matriculation and usable by computer, but for any printed matters, books, articles of newspapers or of periodicals, monographs or annual reports of companies and so on. What is worth to be published is worth to be retrieved, and that means to-day worth to receive a bibliographical address convenient for computers. The support of institutions like the A. S. T. M., or the Library of Congress, or the IFLA is not enough: without international agreements, and with out the help of laws specifying the duty of editors in the drawing up of bibliographical addresses, we will go on working in disorder and encumbrance, for our computers are not built to swallow addresses drawn up minutely by hand even under respect of bibliotheconomy rules. New structures are compulsory for efficiency of new tools.

That problem, as an answer to disencumbrance, is now the very urgent one, but one does not feel that it is yet taken under consideration as it obviously earns.

Address of the specialist

«A pocket memorandum-book, patiently set up during years, is a very humble tool which, if it held the ADDRESSES of the more valuable sources, may be much more efficient than renowned libraries» [7]. These «sources» can be the addresses of the more qualified specialists, or the names of knowledged people who may give pertinent advices, such as the bibliographical address of the more suitable article, or the name of the more qualified information center. The power of big companies lies partly on their relationships with many specialists, which means to possess a network of friends leading to information sources. At disposal for American citizens the National Technical Referral Center adjoined to the Library of Congress fills the mission to point out the searcher to the more qualified sources or specialists. But we do not think that attempts of that kind are already known at international level.

Mecanographical listings or retrievals by computer can ease such detection of these «sources» or of these «specialists». That requests to define «sources» and «specialists» by their profiles. Once more we face the same problem: we ought to organize the standardization of ADDRESSES so that we can consult listings or computers more freely and efficiently.

Address of depository

Each year we are more and more aware of published documents by abstracts of the secondary literature. The main justifications of that practice are obviously a faster choosing and the ability of grouping documents under documentary profiles. A standard system of «need groups» profiles would ease considerably the «feeding» of each information center with the material which is necessary for its activity.

But it is not enough to organize the secondary literature for a more exhaustive and faster retrieval of answers. We ought to find where original texts are at disposal. That requests listings establishing a correspondence between bibliographical addresses and the addresses of the known depositories. The absence of a standardized and mecanographical system of addresses means more manual work or, more obviously, no listing at all. One may understand why such listings are still to-day lacking. The Library of Congress attempts to update such listing for the deposit of American books in all the national libraries. The matriculation of American books eased certainly the drawing up of such index. The Belgian Bibliothèque Royale, Brussels, made a catalogue of that kind for the deposit of any periodicals in public libraries of the country.

One realizes easily the meaning of standardized addresses for the making of such an index if these addresses are convenient for updating by computer.

Address of the user

For the same reasons as above, it is useful that users requesting for information retrieval may be defined themselves for the equipment by their own mecanographical addresses. That will ease the use of automatic equipment like microfilms duplication thanks to automatic reading of the name and address of the application.

It is likely that a larger training with mecanographical addresses will lead to discover that other ones may also be useful, for instance addresses of editors, or addresses of booksellers. ADDRESSES of a mecanographical kind are the starting point for disencumbrance of communication if of general acceptance.

Among the five addresses thanks to which we may hope a more flowing communication, TWO are more characteristic of information activity; of course they are the intellectual and the bibliographical ones. The THREE others are much more similar to simple mailing addresses. They will be more easy to treat, but they have also their role in reaching the result. In order that these last addresses may be convenient for the computer and for mailing, they could be studied in the line of the ZIP CODE. There is a being interest to make listings and mailings easier with the same addresses.

Decongestion of communication is urgent. That requests an answer at the level of the community. What we ought to realize, we repeat here once more what was exposed in chapter 1, is to organize for the benefit of anybody the privileged position of the intellectual who is the owner of a private library. More, the use of the computer will introduce better infallibility than the human memory could.

5. Present situation and projection towards the future: comparison

Communication of information changed several times in size during last decades. In the private library of the intellectual the communication was inward. The information center enlarged the circle but still in a defined way with limits in geographical extension and in intellectual profile. The information center appears as an autonomous entity. Slowly, as the customers' needs extended, the center had to organize a better one-another-help with the exterior. But the feudality of «systems» puts the brakes on harmonious external relations. We cannot meet the rational use of modern equipments as the computer, for we never care with standardization of communication addresses. We need now to enlarge the circle to international relationships, but our internal and local habits imperil communication in such a large network.

It is useful to describe such situations in the each day living of the information center. That comparison is to be made between the present state of feudalism of systems and a future for which, as we may hope, the human beings would have organized communication thanks to standardized ADDRESSES. Such a list of examples may be lengthened but the four following ones show clearly the way.

— FIRST. The searcher himself is alone to be qualified for choosing what may give good answers to his information needs. But that may compel him to read hundreds or even thousands of papers before making a good find. That situation leads him to question the more suitable information centers. But he knows nothing about the retrieval system used by these centers. He ought to explain with many details what he is searching for, and in which context his request has been formulated. And he has to give these explanations to somebody who knows nothing or very little about his own activity and target, but before all somebody whose qualification is to be acknowledged with the retrieval system of the center. Between the searcher and the information scientist of the center, communication can be nothing other than an uncomfortable one. The consequence may be — and often it is so — an unsuitable answer, an answer that gives information but not what the French word «renseignement» means. A «renseignement» is discovered by the user himself and not by a too precise answer given by information centers. — The situation will be different if a generic system of «need groups» categories allowed the searcher to define his needs not more in specialised words, but inside a general frame. By questioning with a «need group» category, the answer will have less chance to escape the «renseignement». Need group answers give information about and around the need and the ability of choosing remains the one of the searcher himself: he opens the dialogue with the most suitable papers. The generic approach respects the personal behaviour of the searcher, while the transit through the information scientist of the center introduces a break in the communication.

— SECOND. To-day the bulk of the published papers is so heavy that the center loses any ability to make an exhaustive choice or treatment. Transit by abstract journals is compulsory. But abstract journals are redundant, and their indexes are built under rules of non-compatible retrieval systems. To compare the abstracts given by different journals and to make the best selection for more suitable answers may be — and is indeed — a tiring puzzle. The situation will become quite easier with standardized generic approach and bibliographical addresses. The abstract journals will be sold and updated on microcards, while the generic indexation will be received on tape for the computer. The generic approach will not only respect the behaviour of the searcher, but the whole collection of abstract journals at disposal becomes easy to be consulted by a single retrieval request. More, the different abstracts published for one paper will be identified by the computer thanks to the bibliographical address, and one may choose the best ones. Standardized addresses organize implicitly the communication between abstract journals.

— THIRD. To discover the nearest deposit library for articles which are recommended by computer retrieval may be slow or impracticable. We lack these listings which give relations between bibliographical address and addresses of depository. By a general agreement, the computer might give answers by standard bibliographical addresses, and lists may be used as input for another programme which will give the concerned addresses of deposit. Papers printed as answer by the computer might be used as order sent to a microfilm library. Standardized addresses organize implicitly the coupling between the computer and the microfilm library which sells hard copies or duplication of microfilms.

— FOURTH. Private matriculation numbers are used by any information centers for internal communication. They permit control of the communication between the document, the abstract, the indexing and the input on equipment (computer or others). But they are of no use outside the center. More, such a matriculation system is not at all a bibliographical address and remains of no help to preserve against making several indexings and inputs for the same document. — Things will be quite different with a standard bibliographical address which may be used for retrieval by computer. Then a one-another-help between centers will become faster. Detection of duplicate indexing will be easy by computer. The habit of private matriculation numbers will become useless, and information centers will be able to file their abstracts in the alphanumerical order of the bibliographical address, the same order for any center. That last possibility will be, as we may think, the powerful tool of communication between centers. In 1965 already [4], the author detailed the numerous advantages of the bibliographical address given by care of the editor.

Without a sane standardization of addresses, information will not easily reach the stage of a trade or of an industry. We do not know any other structure that can make communication faster, safer and more efficient. Thanks to agreements on communication, which means some restriction of autonomy, efficiency may become very large. The proposed structures may open the doors to new procedures for progress. One of the more efficient will probably be the impulse given to the literature of rewriting, of surveying and of state-of-the-art reports. Progressively the mission of the specialized information center, which is to set up a «da-

* * *

ta base», will extend till the «data bank» mission. Another consequence will be, for any company, that the industrial information center will be relieved of a part of the burden of updating the collections, and will become more able to fulfil its real role in the company: the drawing up of surveys which bring an efficient introduction to the decision. We may pronosticate that communication of information, thanks to these new structures, will change in size another time, allowing from that moment a more rational use of computers and equipment.

Information will become «moving» and will find better opportunities to contact qualified and cognizant people, and to fructify. One will be more able to disseminate the information and to use it in the outline of the definition given by François Perroux [8]:

«L'information est un bien de connaissance qui s'achète et se vend, qui est produit et distribué, qui comporte un coût et procure un rendement, qui est astreint à des règles économiques en quelque organisation que ce soit»

> (Information is a good of knowledge which one buys and sells, which is produced and dealt which requires a cost and raises an output, which is tied to economical rules in any organization whatever it may be)

We may think that a better standardization of communication thanks to addresses of general acceptance will ease the development of a more harmonious network of specialised information centers, with the objective of «reviewing» knowledge, which is the most significant proposal of the Weinberg Panel report [9]. Such reviewing of a field of knowledge opens the door to the more suitable bibliographical references: «If you like to dig deeper see so and so» [10].

These information studies are expensive, the more so because they have to be drawn up by very qualified people. The only possibility to lessen the relative cost of such reviewing studies is to sell them to more customers. That is certainly an important justification for better communication. Information reviewing will become a business success as much as a procedure for progress and education. Trade of information claims for its place aside trade of books and periodicals.

6. Awareness by each of us

Information of knowledge even to-day is probably the single human activity «of communication» which does not succeed in leaving the artisanal stage for the one of an industry or a trade. Information activities are unable to rise above the encumbrance level, for they are unable to amend by their own.

It is likely that the most important cause is a poor understanding of conditions of efficiency for computer and other modern equipment.

It is not seldom but however surprising, to meet people confusing information itself with the document that is nothing else than the holder of information, or people confusing the «message» and its «transmission». Such confusion, enclosed in our habits with the big role of libraries in the information field, carries its consequences:

One had inclination, and one has still inclination, to entrust information treatment to those who are traditionally responsible of documents with the effect to give to information the centralized structure which is the one convenient for documents treatment. Slowly but surely awareness stretches around the world that information treatment requires a close contact with specialists of the concerned field of knowledge, that means an unavoidable decentralization towards these specialists. From that moment the dynamic part of the treatment cannot come from the document itself, but from the huge specialists network disseminated in any active place for progress, for practical application, or for research. One understands also that the bidder of information can in dissimilar circumstances become the creator or the seller of information. Everything is moving inside a network where dynamism cannot be kept up unless communications of a high standard exist. We must recognize that such dynamism which is a result of communication, is really running in the opposite direction with information treatments centralized in more and more powerful knots far away from the specialist worker. Such duality of outlines may imperil the result.

As consequence of the above considerations, the solution appears more clearly: it lays on a rational organization of communication and on standardized addresses which will enlarge the use of computers and equipment.

It is a reform of the structure. Perhaps some responsibles presently in charge at a place of the information chain will put the brakes if they do not understand that the best is to go with progress.

The mission of communication appears as a main mission for the community. It ought to be ruled between governments by taskforces or delegations responsible for the solution to be chosen, but doing so in the outline of future and not in the frame of present obsolete structures.

With regard to the position of each of us, the best we can individually take is to decide in favour of the dynamic conception of communication. Each of us must be aware that the reform will have a heavy impact on the future, and one can be warrant to have its efforts to adapt himself, rewarded by the result.

It should be desirable that a group of people who are well sensibilized to that problem of structure will set up a design convenient as starting point for an international move via the national governments.

REFERENCES

1. «L'homme encombré», collection «Prospective», 15. Paris, Presses Universitaires de France, 1969
2. — «Scientific and technical communication». A pressing national problem and recommendations for its solution. Report... X111, 322 pp.
   — Synopsis of the Report. 30 pp. Washington, National Academy of Sciences
   — National Academy of Engineering, 1969
3. Excerpta Medica Foundation, 364, Herengracht, Amsterdam.
4. J. Halkin, L'adaptation du document à sa fonction. «Bulletin de l'A. I. D.», 1966, January
5. CODEN for periodical titles. ASTM special technical publication № 329, 1 st and 2nd eds. Philadelphia, Pa., American Society for Testing and Materials.
6. C. Vernimb, Research and Information. «Euratom Bulletin», 1964, 3, № 3 (EUBU 3—19).
7. J. E. J. Halkin, L'innovation et la communication de l'information. In: «Etudes et Documents du CNBOS», № 292, 1967, p. 10.
8. F. Perroux, «Economie du XXème siècle».
9. «Science, government and information». Washington, U. S. Government Printing Office, 1963
10. Form used by the American periodical «International Science and Technology».

D. B. Baker (USA), **P. V. Parkins** (USA)
and **J. Poyen** (France)

THE FUTURE OF ACCESS (ABSTRACTING AND INDEXING) SERVICES

> «There is a growing realization that man's future may be literally what he chooses to make it, and that the ranges of choice and the degree of conscious control which he may exercise in determining his future are unprecedented.»
> John McHale

1. History

Over the centuries scholars and research workers, as well as editors and librarians, have recognized the great importance of bibliographic control of the literature via indexes, catalogues, lexicons, registers, abstracts, reviews, digests, and the like. The main purpose of these tools has always been to record, to organize, to alert, to retrieve, and in modern terms — to access — the information and knowledge of the universe.

Indexes have been relatively easy to make and were one of the first locating devices to be started by classical and Biblical authorities many centuries ago. The first volume (1545) of Conrad Gesner's enormous «Bibliotheca universalis» was more of a classification system, a comprehensive international dictionary, than an index, but his second and third volumes, «Pandectae» and «Partitiones» were subject indexes. Indexes were well established when the first periodicals appeared on the scene. It was a logical and inevitable extension of the use of indexes to apply them to the first journal. The first volume of the «Philosophical Thansactions», Royal Society of London (1666), was issued with a subject index.

The review journals in the late 1700s may be regarded as the intermediate between the collective periodical indexes and the abstract journals. In 1771, in the review of «Medicinische — Chirurgische Bibliothek» in Germany, it is stated, «a number of publications I have myself seen and abstracted. Most of the abstracts, however, are from the best journals, and frequently literally transcribed, even when I had the book in hand and had read them through since this is in accordance with my plan.» While it is foolhardy to claim the title of «first» for any bibliographic event, certainly the «Pharmaceutisches Zentralblatt» (1830) a forerunner of «Chemisches Zentralblatt», was one of the earliest abstracting and indexing journals as we know them today.

In the twentieth century, abstracting and indexing services increased rapidly — some broad and some narrow in the subject scope and in the range of journals covered. By 1930, most disciplines in science and technology and in the more technically advanced nations had some secondary services. There were numerous studies between 1930 and 1960 which report the extensiveness of the secondary services in coverage of the periodical literature. It was not until 1963 that «A Comprehensive Guide to the World's Abstracting and Indexing Services in Science and Technology» was published by the U. S. A. National Federation of Science Abstracting and Indexing Services. This guide classifies and records the details of some 1855 secondary services in 40 countries. A FID Guide (second edition, 1969) of abstracting services alone identifies 1300 world-wide in Science and Technology (Vol. 1) and 200 in Social Sciences and Humanities (Vol. 2). A joint revision of the NFSAIS and FID guides is proposed for 1973—4 and estimates show an increase of secondary services to be 1.5% per year.

However, in 1947, the United Nations Educational and Scientific Organization (UNESCO) undertook a program «to facilitate the improvement of scientific documentation and abstracting» in accordance with its charter «to maintain, increase and diffuse knowledge — by initiating methods of international cooperation calculated to give the peoples of all countries 'access' to the printed and published materials produced by any of them.» One of the first international conferences on contemporary scientific information was held by the Royal Society in London in July, 1948, when two sections of meetings were held on improvements in publishing and distributing of abstracting and indexing services. Resulting from the UNESCO program and the Royal Society Conference, an Abstracting Board was formed under the auspices of the International Council of Scientific Unions (ICSU AB) in 1952. The purpose of ICSU AB at the beginning and continuing today, is «to organize and promote the international exchange and publication of information in sciece and technology.» ICSU AB collaborates with other international organizations with related programs, for example, in regard to the interaction between scientific disciplines and the use of modern methods of information processing: UNESCO, FID, IFLA, IFIPS, International Organization for Standardization «ISO), and ICSU Committee on Data for Science and Technology (CODATA).

2. Current Status

2a. Growth of Material

In «Little Science, Big Science», published in 1963 by Derek de Solla Price, he points out that the growth rate of printed abstracts has stabilized at a level which effects a cumulative doubling in each 15 years. This generalization can be disputed in particular disciplines, but it is indicative of a fairly constant growth rate over a long period of time. And if true, it means that between now and 1986 as much material will be published as has so far appeared in print during the whole of scientific publishing.

Although Professor Price was discussing the growth of published abstracts, this of course reflects a rate of growth of similar dimension in the primary literature. Indeed some discipline-based access services are aware that because of economic and other constraints, they have not always been able to keep pace with coverage of the primary literature to the extent that they would have wished.

Throughout at least the last 25 years there have been regular predictions that the growth rate of the primary literature must decline, and that if this does not happen naturally, then some artificial restraint at the source would need to be applied. To date this has not happened, either naturally or by artificial means, and the secondary services would be poorly advised to base their future planning on such an assumption.

2b. Improvement in Tools

The massive volume of printed material has introduced problems not only for the processors of secondary information but also for the users. Conventional methods of production of abstracts have been stretched to the limit, as services began to handle a quarter of a million items per year or more. Equally strained are the information search methods of users, for no scientist can hope to keep pace with all of the literature in his full discipline and, indeed, he may face very serious problems in keeping up with the literature in his own narrow field of specialization.

Fortunately, the becoming available of electronic computers has in this, as in so many other instances, been helping both the producer and the user. It is wrong to believe that the availability of this modern tool provides the solution to all problems, but computers are certainly making much possible which could not otherwise have been done. Firstly, it offers the opportunity for massive storage of information in relatively small areas. It also provides an efficient and rapid searching tool which means that retrospective searches of the literature which would have required weeks or months by hand can now be accomplished within a few minutes. However, it does not as yet function as the intellectual interface between the questioner and the system; the conversion of a question into language suitable for searching the computer file is still a very special job. Some experiments on direct retrieval by users have been successful but these have so far been limited to relatively small files. The ability of every scientist of search a massive file is part of the planning of most large systems; it is not yet accomplished routinely by any means. It is one of the misfortunes of scientific language that scientists do not always use the same words to mean the same thing; otherwise the retrieval of information from a machine-based file would be a much simpler matter.

Computers also offer facilities for selection from a main file which can be searched in any way needed and produced as magnetic tape output for driving photocomposing machines which in turn can typeset the published material much more quickly and efficiently than conventional printing machines. They also enable current awareness searches to be conducted on material shortly after it enters the file. This means that a scientist having a specific problem may be constantly alerted to anything new published in his field. Nothing in his speciality may be newly published and if so he will know this also; and no matter how infrequent publication in his interest area may be, he can be assured of its being brought to his attention.

All of the foregoing which at the moment seems so advanced and ambitious will doubtless in a few years seem quite primitive. New methods of recording scientific information different from the conventional ones are being developed and tested, and will be adopted. Methods aimed at machine-based services will be used and many of the problems will disappear. It is the almost impossible task of attempting to convert a system which has employed the same techniques for centuries to one using modern tools which introduces most of the difficulties. Complete rethink and redesign will be necessary.

In addition to the application of computer communication techniques, changing facilities for the rapid transfer of material from one place in the world to another, e. g., by communication satellite means that national barriers in information transfer can be broken down and opportunities for true international cooperation become more readily available.

The different languages still present a problem not only of translation but of understanding what the same words mean to different people. An developments toward a language of science would be of tremendous benefit in the transfer of scientific information.

3. Future

3a. Relationship between Primary and Secondary Publications

For the purpose of this discussion the information processor is defined as an organization that provides primary and/or secondary publications and services. The primary publications are the true repository and record of the original, basic data and information in discrete, printed packages. Secondary services provide organized, subject-oriented access to newly available information from a specified range of primary publications — serials, patents, and reports — and/or to the accumulated store of information available in such documents. Secondary services do not replace the primary record; abstracts and indexes are not surrogates for the primary documents. While some primary publications carry in each issue a section, limited in scope, of abstracts from other primary journals, the large, major secondary services have developed as separate functions with separate operations in the past two centuries, with very little coordination between these primary and secondary processors.

In the traditional publishing system, each information processor has independently established its scope of coverage, editorial practices, organization of data, timeliness of reporting and publication formats. Cooperation among primary and secondary processors has been severely limited in the traditional system because of the manual transcription processes which are required to transfer information from one operation to another and which have prohibited such cooperation from being economically feasible.

In the developing computer-based information-handling systems, however, the functions and component parts of the processors must be increasingly automated and coordinated. Such automation requires fully defined self-consistency within each processor's operations and highly precise recording of the information handled within each component. Self-consistency does not necessarily imply identical practices for all processors. However, exchange of machine-language records between two or more processors does require that each processor operate at the same level of detail and that each operation follow its own defined practices consistently.

Many experiments have been undertaken by the processors throughout the past decade to further compact the information, shorten the time in process, decrease duplication of intellectual effort, reduce costs, increase the number of papers or abstracts handled, increase utility of indexing, and increase reliability of the recorded information. In addition, a majority of primary publishers are now providing abstracts along with the full paper. Most are using more informative titles than in the past. Some are using codes to identify uniquely their journals as well as each paper within the journal. Further, some are experimenting with the generation of appropriate index terms, diagrams, formulas, and nomenclature to be used jointly with secondary services.

The growing practice among primary publishers of storing part of the author's data in a depository in lieu of publishing it presents no serious problem to users so long as the deposited material is made available to the secondary services and is widely and easily accessible to libraries and individual users. In recent years, some primary publishers also have started single-article announcement services which permit users to order relevant articles on the basis of titles and/or concise summaries.

Hopefully, the experiments and plans being made today by information processors will lead to sound cooperative systems of the future. These plans are designed to uncover and solve the problems of primary and secondary processors working together in the development of a total information processing system. As the literature has grown and as editorial and production costs have escalated, it has become mandatory to eliminate the duplication of intellectual effort which goes into the analysis, editorial, processing, and indexing of every scientific document — steps at present largely duplicated by primary and secondary processors.

The experience gained to date has enabled primary and secondary services to design and start building more efficient systems to interlink the processing steps. It seems likely that such interlinkage of operations will be further extended and broadened in the decade of the seventies both on a national and international basis.

In the meantime, some other steps can and will be undertaken as the computer-graphic processing equipment and computer-based information-handling procedures are installed by more and more information processors. It is being suggested more each year, as the pressures grow, that the full primary record of the data and information can be deposited in microform, and perhaps even in the digital form eventually, throughout the world in the large central libraries. The alerting and retrieval access functions by some combined primary and secondary services will perhaps be well along by 1980. These processes will necessarily be well coordinated with activities of the information centers that currently are being established worldwide for evaluating or selectively disseminating information as well as with the libraries which provide the original document handling and back-up services for the full record.

These trends and advancements will certainly offer the users the following advantages: increased reliability, improved access, more timely awareness, improved responsiveness, and reduced costs and better overall economic performance of the system.

3b. New Trends for Secondary Services

The power and flexibility of the new information storage, retrieval, and dissemination techniques that have been developed or now are being developed are having a tremendous effect on abstracting and indexing methodology. The major secondary services are undergoing rapid change and acquiring new, more extensive information-processing and disseminnating capabilities. The use of the computer and graphic devices to repackage information has proven to be a most important development of the 1960's in its effect on the ability of the abstracting and indexing services to better meet users' requirements. The variety of possible forms of output from these systems is virtually unlimited.

In these computer-based systems, information selected in a single intellectual analysis of the source documents, an analysis which combines both abstracting and indexing, is put into a unified machine-manipulatable store through a single keyboarding operation. Then from the unified bank of information, material appropriate for special-subject alerting and retrieval can be drawn, largely by computer program.

The large, comprehensive, printed issues and volumes of abstracts and indexes continue at present to be the most used — and, therefore, best supported — products of the secondary services. However, the distribution and use of microform and magnetic tape versions of the comprehensive volumes as well as selected subsets of data are increasing each year. Some secondary services, in addition, are attempting to provide specialized bibliographic and demand searches directly to the user community on a broad basis. Other secondary information processors are providing their data bases in a gross or semi-refined form to information dissemination and/or evaluation centers which further refine the information, often combining it with other data bases, and provide selective dissemination to individual users or a particular community of users. The number of these information centers with computer processing capability is increasing each year. Over 100 such centers were in operation in the United States alone by the end of 1970. Many nations have established or are in the process of establishing one or more such information centers to serve the various scientific and technical disciplines; the number of centers depends upon the size and needs of the nation. In some instances, several nations are cooperating in the support of a single center.

It is too early to predict the exact role of the secondary information processors in the future — whether they will continue to provide a full range of publications and services directly to the information users or whether a substantial portion of their services will be provided through intermediaries such as information centers. The options are many, but coordination at all levels will become increasingly important as the use of computing systems grows.

3c. Future of Publications

Let us start by saying that for the foreseeable future, some form of the printed media is here to stay. It is hard to conceive, so long as the human race continues to be so mobile, that anything can replace the portable printed medium. The facsimile delivered to the home, the office, or the laboratory is only a partial substitute, though it will undoubtedly become a reality.

There will be an increasing opportunity to have electronic or graphic presentations of information in the home or office via terminals, supplemented by printed background material in hard-copy form. The major obstacles to this form of communication today are the high cost and the lack of satisfactory terminal and electro-mechanical output equipment. It is doubtful that the technical problems will be overcome and he costs reduced to a reasonable level before the end of this decade.

In the future, people will read more because they will have more time to read and they will need to learn more, although their reading will probably be in a context completely different from today. These reading patterns are already changing, with the integration of electronic communication and printed and graphic material with various audio visual presentations.

Undoubtedly, these changes in reading patterns, and a trade-off between the printed publications and the electronic communications, will continue. The information processors will need to keep close to research and changes in reading habits of the users to adapt the content and form of their output.

3d. *Future of Tools*

3dl. Hardware

As used here, hardware is defined as mechanisms of all types used as tools in «access service» processing. Included are computers, photocomposers, and input devices, as well as the software necessary to control and operate these hardware devices. There is no question that, electronically, hardware is getting more powerful, faster, cheaper, and smaller, much as has been projected for many years. All recent technology forecasts continue to claim these trends, and the advertisements of new equipment are consistent with the forecasts.

The ability of routinely available input and output hardware to handle letters, numerals and other typographic characters now has expanded from the early 48 to 64-character sets (upper case only) to the 95 characters (including upper and lower case) of the American Standard Code for Information Interchange (ASCII). This does not yet satisfy the full range of character requirements for many access services, which are surmounting the problem through specially designed printer and photocomposition units. While the basic electronics capability for expanded character sets exists in various cathode-ray-tube devices and photocomposers, a mass market for such devices is needed to encourage additional development and reasonable costs.

Of particular concern to information processors is the development of computing equipment memory with respect both to size and cost, but the capability is greatly improving each year. A more difficult task is ogranizing the information to be stored and providing suitable supporting software. There is much work being done in this area now. This is a key problem facing all access services in this decade.

Input mechanisms are developing very well. Graphic input devices are in their early development stages. Voice input devices will not be technologically or economically feasible for large-volume input in the near future. The most important advance for access services in the short range future will be to capture as much of the input as possible as part of the primary publication process.

Microform production and handling mechanisms comprise another important form of hardware for the information processor. The role of this equipment in information handling, whether as processing aid or distribution medium, is still not major. However, as judged from the latest conferences, the variety, capabilities, and prices of microform equipment are improving very rapidly.

3d2. Automatic Indexing and Abstracting

Text processing by machines to provide functions equivalent to abstracting, indexing and other forms of linguistic analysis has long been a subject of research. Only a little over a decade, however, have serious efforts been directed to the use of machines for «automatic indexing». At this time, «automatic indexing» in the form of KWIC and coordinate indexing is widely used. These indexes provide users with enough access points to the literature to be somewhat acceptable, and at the same time they are economical to produce. However, these methods do not meet the criteria of excellence and completeness or compare favorably with some index methods currently carried out by humans. Thus, the problem is how to improve the relationship between performance and economic efficiency so that one can develop computer-based indexing systems with the degree of sophistication needed to satisfy user requirements.

The difficulty of the process of automatically selecting index headings and the problem of transforming source index material to a format suitable for modification under these headings have delayed the use of automaticindexing. Present indications are that computer-assisted indexing techniques may be more feasible than complete automatic indexing directly from original documents. It seems natural that if we let the human analyst retain the task of selecting meaningful headings from titles and sentences suitable for computer processing, the tedious, time-consuming, and difficult procedures of articulation, permutation, and the like could be performed by the computer.

The preparation of an abstract from a full report can be a difficult process, especially if one demands from the reviewer a critical resume or an evaluation of the reported work. On the other hand, preparing indicative information on what an original document is about is a much simpler operation which can in many cases satisfy the needs of the user for information. It has been demonstrated that machines, with the help of a few rules of semantic and linguistic origin, can extract sentences from originals just as humans do, although the indicative abstracts produced by extracting key sentences should probably be called «extracts» rather than abstracts. The algorithms developed to extract these phrases and sentences from documents are based on approximations of human procedures involved in the process, for instance, of underlining some important passages from a book.

Such computer-produced extracts have the same organization and vocabulary as the original, and for the extraction process to be effective, the original documents need to be written in an expository style and have a logical organization. These are generally prerequisites for publication of scientific papers, which makes automatic extracting widely applicable with little, if any, pre-editing.

The lack of ready availability of primary documents in machine-manipulatable form currently limits the application of automated extracting methods, but with increased use of computerized primary publication processes, the input material will become available. Lack of inter-sentence coherence is another problem of automatic extracting. This could be solved by human editing; the process of producing extracts with the aid of a computer does not need to be completely automated. Finally, semantic and syntactic criteria for algorithmic extraction still need some improvement so that more reliable extracts with smoother transitions can be produced.

Thus, the present state of the art indicates that automatic or semi-automatic indicative extracting of scientific articles may be possible in the near future. It will come at first with the goal of reducing, by mechanical means, the size of textual materials so that readers may be advised without much effort whether they should read the entire original document or not.

On the other hand, informative and critical abstracts will continue to demand sophisticated semantic and syntactic rules for content analysis and deep structure understanding, all of which are highly intellectual processes quite difficult to approximate by machines. Such a process is as complex and difficult as automatic, natural language translation and speech recognition, which are still at an academic research stage. Many problems remain unsolved, including the need for further development in linguistic theory significant to automated language processing, before such processes can be considered by non-academic institutions.

3d3. On-Line Users

On-line access to information, which persists as a dream of most information users, is today a reality only for a favored few and for a limited scope of material. It is, however, a correct vision of the future for many, with the only unresolved question being «when?». The timing will depend upon many factors. Some users currently have access to a major computer system that can be used for on-line information activities on small files, and this capability can be extended to large files as large capacity direct-access memories become less expensive. However, for most users, on-line searching will depend upon the development of cheap, high-capacity communications and terminal facilities to provide access to files maintained remotely. These are technically feasible at this time, much as space travel was scientifically feasible in the early 1960's, but there must be a sufficient incentive to justify the development costs.

As important as the world scientific and technical information complex is, it does not offer a large potential market or revenue potential as compared with that for telephone communication or television transmission. Thus, while the technological problems that must be solved to make widespread use of on-line systems a reality are no more difficult than those that have been solved in some other area of electronic communications in recent years, the cost of solving them must be borne by a much narrower base of support, and the social and economic incentives for solving them do not seem to be sufficiently strong at this time. Moreover, not everyone has the aptitude or training necessary to make effective use of on-line systems. If these systems can be inexpensive enough, this probably is of no particular consequence, but if a large expense must be justified, it will be necessary to show a significant improvement over other methods of access. The on-line retrieval system is really no more than a vehicle that the user pilots through information space, and the effectiveness of the results depends upon the pilot's skill as much as the inherent capability and organization of the system and the files within it. There is a basic difference of approach between the «continuous-absorption-to-be-prepared-for-anything» method of using information, which is traditional current awareness, and the «request-what-is-needed-on-demand-and-don't-worry-about-irrelevancies» method. The difference is the strategy of use interactive systems.

3d4. Digital and Satellite Communication

At some point in the future (as yet an unpredictable time) digital communication facilities, including satellites, will become an electronic marketplace for information and computation services. Conceivably, such could provide the «transportation system» which enables geographically remote specialized services to be brought to the information consumer. When such a situation develops, the «access services» will find it desirable to enter that marketplace, either directly, of indirectly through information distribution centers.

One can imagine this type of future and it can be designed. However, it is not yet time for the access services to take any special steps in that direction. There is enough to do now in routinizing and restructuring present processes and services. If this restructuring is well done, it will be applicable for extension to a communications-oriented world, if and when that world develops. If it does develop, it will not be uniform worldwide because of the uneven state of technology advance in various countries. Therefore, services with worldwide interests will have to operate in a mixed marketplace.

4. New Trends in Coverage

Traditionally, secondary information services have developed by discipline and by language. In some countries such as France and the U.S.S.R virtually all fields of science and technology are encompassed by one centralized information service; but in others such as Germany, the U. K. and the U.S.A., separate services, usually based on professional societies of scientists have been created as the demands for them have arisen.

This means that, for example, there are separate services in physics in English, French, German and Russian languages with extensive duplication of coverage. Although this duplication in coverage may in some cases be desirable, the procuring and handling of the same material many times in as many different places may be wasteful.

Another more recent development has been that of mission-oriented access services, devoted to the support of such fields as atomic energy and oceanography, where information services are required which span two or more disciplines as necessary to cover a particular project. These services may be national in scope or international. «Nuclear Science Abstracts» offers an appropriate example of the former; in the international field, INIS is a new and good illustration. Many of today's problems, because of their breadth of interest, are interdisciplinary in nature. Thus in the future we can expect greater emphasis on multidisciplinary systems, It is fortunate that now there is a relatively new and growing recognition that, in proper combination, many of the long standing discipline-based services are capable of providing the subject coverage needed to establish a mission-oriented service. For whatever reason, in the past we have witnessed a tendency to create anew, to begin all over again without taking advantage of information sources already available. To some extent the discipline-based services are themselves to blame, for they have tended to pay insufficient attention to developing compatibility among the services, and the problems of converting this material so that it can be used in one file have been tremendous. Now, however, discipline-based services are recognizing their responsibilities, are cooperating much more actively, and joining together in many aspects of their future planning. One can state without exaggeration that even now, it should be possible to meet most future requirement for mission-oriented access services from a combination of selections from discipline-based services. Plans should be laid to that end.

It is difficult to predict what multidisciplinary services may be required. It should not be **necessary** to predict, however, since the coverage of all scientific fields is reasonably comprehensive. With relatively slight modifications and additions, with where advisable some changes in emphasis, it should be possible to meet demands as they arise. Considering a current example, research interests in space exploration have in large part shifted to ecology, the latter loosely defined in a variety of ways. Although newly highlighted, ecology is not newly discovered. And preserved in the files of several discipline-based services is a massive store of vital information for those interested in this field. Furthermore, the machinery already exists and is operating to capture all scientific writings of ecologic importance, and to analyze and store these appropriately for future retrieval. Such «stores» can be quickly tapped to meet current demands. It is in this direction that the development of mission-oriented services must move, if it is to be possible to bring adequate information promptly to bear on the world·s complex changing, multidisciplinary problems.

5. Cooperation

5a. *Present Situation and Needs for Cooperation*

Before looking at the future of access services from the cooperation point of view, let us look at the present situation and identify some of the needs for cooperation.

First, for a truism: the amount of primary literature has been growing at an exponential rate for several centuries. Thus, it is becoming more and more expensive to acquire, analyze, process and disseminate this information to users, pertaining to a given field of knowledge, whether mission or discipline-oriented. And the increase in cost due to the improvements in the services being rendered is additional. It has become obvious to most specialists in the field of information transfer, and in particular to most managers of access services, that they will not be able to face this growth situation forever, because of financial pressures. They will not be able to afford to maintain high standards of completeness, timeliness, and quality in the future, as in the past.

Then, what are the alternatives? One would be to abandon completeness; that is, keep coverage more or less constant, or at least at a growth slower than the growth of the literature of the field. This means that the percentage of the field covered by the access service would gradually decrease. Such a policy would lead to a proliferation of access services; new ones would be created to cover the gaps left. Also, this proliferation would not help meet the users' needs across languages and disciplines. On the contrary, it would force the user to look at numerous access services to find the information desired.

Another alternative would be to abandon or decrease timeliness or quality. Here again this would not help the users and would lead to chaous eventually.

Another possibility, so often discussed, is to take measures to decrease or to limit more rigorously the amount of primary literature issued. It is generally agreed that some steps could be taken in this area which would help. But the scientific and technical population of the world is increasing, as is the level of civilization in this scientific and technical era, as well — and, therefore it can be assumed that the literature will follow this growth in one form or another.

Such a situation would seem desperate if there was not another factor, paramount in the overall system, e. g. duplication of efforts.

Until recent years, the access services have grown up independent of one another. This has led to a great duplication of efforts within nations and among nations. The same material is acquired, analyzed, and processed in so many different places that it is impossible to obtain factual data on how many times a given paper is abstracted and indexed. The total amount of money thus spent all over the world on duplicating one another's efforts is enormous. This sum is considered to be many times greater than the amount necessary to render the best possible, coordinated services in all fields of knowledge and to develop the highly sophisticated tools and modern systems which would increase measureably the efficiency and effectiveness of such services.

Also, the independent growth of access services leads to a situation where some fields are covered several times, and the users in these fields have their choice of services, while other fields are covered only very partialy or not even at all.

The most obvious and, perhaps only, remedy to improve this overall situation is to develop cooperation and coordination among the access services, as well as with the primary pulbications, libraries, information centers, and other organizations involved in the information transfer chain.

The idea of a centralized system is not new. One can go back to an International Catalogue Conference organized by the Royal Society and held in London in 1896, where the concept of a world center for information was recommended which would cover all fields of knowledge and render all the possible services in all disciplines and all languages. As attractive as such a concept can be in philosophy, it is obvious that it is totally unfeasible today. It would lead to ask the existing services to stop all their activities in order to be replaced by a mammouth center which would be created from scratch. One can easily contemplate the size and complexity of the technical, financial and political problems which would be created.

This is why the concept of a world center has now been replaced by the concept of a world network of existing and future services. Within such a world-wide network, the services would operate in a cooperative manner to avoid the duplication of efforts and to identify the gaps and cover them according to a mutually agreed plan.

5b. *Bilateral Activities and Associated Risks*

The need for cooperation is so urgent that some access services have already started to develop limited networks.

Chemical Abstracts Service, International Atomic Energy Agency (IAEA) with INIS, MEDLARS, VINITI with COMECON countries, etc., have started to cooperate with other services, usually covering the same fields, and to develop information sub-networks, which can be described in an over-simplified way, as follows:

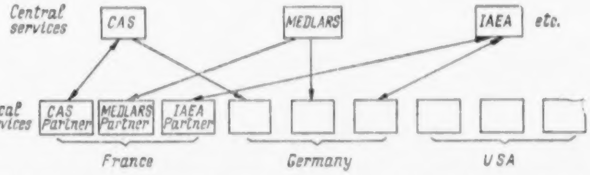
> In one country or region, agreements are made with local services which are responsible to cover the source literature from that country or region and to send the information to the central processing unit. In return, the local services receive the publications and services from the cenrtal unit for searching purposes to answer the questions for their community of users.

There are major problems which have to be solved before such sub-networks can be put into full operation, such as standards, definitions of coverage, editorial and format policies, etc., so that the information which the central service receives from its local partners is fully compatible with the specifications and requirements of the system.

One problem is that these standards vary from one central service to the other. When a local center is started in France, for example, as an IAEA partner, the French center conforms to IAEA standards. When another center in France becomes the MEDLARS partner, it conforms to MEDLARS standards, and so on.

Although such bilateral and multilateral agreements are steps forward in certain fields and they may improve the immediate situation, they do not achieve a major improvement in an over-all access system for the long-range future.

Let us imagine what the future would be, if the uncoordinated scheme was to be followed. There would exist a significant number of central services each covering a given scientific field (like CAS) or technical mission (like IAEA).



This seems fine until one considers that there is a significant percentage of duplication among the coverage of the central services as well as in the coverage of local services in each country or region.

It is true that there have been some improvements in recent years. The French literature in atomic energy, for instance, is not acquired, scrutinized and processed at the same time by centers in all countries of the world. It is processed once by the French IAEA correspondent, but atomic energy literature may also belong to engineering, physics, biology, etc., so that the total French literature may be processed by the various local French services several times.

As has been described in 4 above, the central services in mission-oriented areas will proliferate, because of the increasing needs for project-oriented information. Thus, the duplication of efforts if uncoordinated in a country or region will also increase and result in much inefficiency and eventual chaos. This is the most compelling force at work in favor of the concept of a world network of existing services which consists of having «a mutually agreed upon, world cooperative plan».

The above is not to say that bilateral cooperative efforts are useless. They help to identify the problems raised by cooperation in practical, operating modes and lead to solutions of some problems and in some areas. From the evolutionary point of view, bilateral arrangements and sub-systems must be developed and encouraged. But what the information world really needs is to develop cooperation among the central processing, access services in the framework of long-range, world-wide plans and goals.

5c. *The World Plans*

5sl. Unisist

In 1966, the President of ICSU proposed to the Director General of UNESCO that the two organizations jointly study the feasibility of a world science information system. A Working Party in January, 1967 established the guidelines for the study and defined any «system» to be considered in the feasibility study in terms of «a flexible network based on the voluntary cooperation of existing and future information services.» Furthermore, «feasibility» as used by UNISIST, was understood mean political and economic feasibility, rather than technical feasibility.

A Central Committee and Working Groups were established. The most germane to access services were the joint UNISIST—ICSU AB Working Group on Bibliographic Descriptions and an Advisory Panel composed of representatives of large operating service most of which are also member services of ICSU AB. In addition, the Central Committee commissioned studies on comparative classification systems in the sciences and another on the feasibility of an international register in machine-readable form of scientific journal titles.

The Central Committee also commissioned Prof. J. C. Gardin of CNRS to draft a final report of the feasibility study. This report is now being prepared for publication. Also, a synopsis or shorter version of the Committee's recommendations is in preparation.

The Director General of UNESCO will convene an Intergovernmental Conference in Paris in October, 1971, to consider the recommendations of the Study Report. It is expected that the Conference, after reviewing the recommendations in full detail, will make recommendations to the Director General of UNESCO relating to their implementation.

The 21 recommendations in the Study Report which have to do with potential information programs are organized in five groups. Group one contains 6 recommendations, calling for the development of tools for systems interconnections. This group contains a concentration of recommendations of concern to the access services. For example, **Recommendation 2** calls for the development of an International Serials Data System. **Recommendation 3** relates to standardization of elements of bibliographic description. **Recommendtaion 4** calls for a renewed attack on the problems of subject specifications in the sciences. **Recommendation 5** concentrates on resolving some of the problems of incompatibility among systems.

The second group is concerned with the support and the strengthening of existing services. **Recommendation 8** follows since it refers specifically to abstracting and indexing services: «Ongoing experiments and programs aimed at increasing the effectiveness of abstracting, indexing, and translation services, geolinguistically disciplinary or mission oriented, should be publicized and encouraged. Particular support should be given to cooperative schemes resulting in an international sharing of the work and product of such service, as essential building blocks of the world wide information network to which UNISIST is dedicated.» The accompanying text refers specifically to ICSU Abstracting Board as the nucleus of an action group to which UNISIST must look for the implementation of these recommendations.

If the Intergovernmental Conference succeeds in its mission as planned, it may be anticipated that the sympathy and understanding of governments will be directed to the support of scientific abstracting and indexing.

It is not possible to discuss here all of the recommendations of the UNISIST Study Report, and most of them have interest for the access services. One can only commend the complete report to the attention of the reader.

5s2. ICSU AB Plans

The ICSU AB is an international scientific organization which federates representatives of ICSU, representatives of international scientific unions belonging to

ICSU (who bring into the ICSU AB the point of view of producers and users of scientific information), and representatives of the major operating access services all over the world.

ICSU AB's role in a world information system consists in being the specialized agency dealing with abstracting and indexing services. To this end the Board has, within the framework of UNISIST, launched a program for a world system for abstracting and indexing services, consisting of three parts, covering; (I) Input, (II) Processing, and (III) Output, respectively.

The concepts of the plan for developing cooperation at the input stage, which follows, have been unanimously adopted by the Board in July, 1970, and endorsed by UNISIST. The Board is now working at designing in detail the studies necessary before implementing Phase I.

The world system for abstracting and indexing services is defined as a conceptual entity; it can be thought of as an aggregate of the major, existing abstracting and indexing services all over the world. Thus, the overall input into this system is the aggregate of the input of all the individual abstracting and indexing services and the coverage of these services. The ideal situation and long-range goal of the input plan will be the complete sharing at the input stage by the member services with no duplication and no areas of incomplete coverage. The ideal situation will also be such that each access service inputs into the system, bibliographic descriptions plus abstracts plus indexing terms which will be standardized and in machine-readable form.

The ICSU AB plans to develop compatibility at the input stage, initially concerns the periodical literature which forms the bulk of material processed by the access services. In earlier studies, the member services have identified and established a relationship with journal productivity as falling into three classes; high (core), medium, and low (fringe) productive journals. It is contemplated that a central redistribution unit could sort the information according to the «field code» (e. g., member services identified for certain classes of material) and send it to the interested abstracting and indexing service. The central redistribution unit may not need to exist in fact, if after the studies it is determined that each member service could more efficiently undertake this role.

The first stage plan defines guidelines for cooperation among the member services in acquisition, selection, and exchange of documents for coverage by the services. This is aimed at eliminating much of the duplication in journal acquisition.

Under this plan, the member services would assume the responsibility of acquiring and selecting articles for coverage from the most productive journals in their discipline. Articles in those journals that are of potential interest to other disciplines would be forwarded in microform, along with the author abstract and a standard computer-readable bibliographic description, to the appropriate other cooperating services, either directly or through a central redistribution unit. Responsibility for acquiring and selecting articles for coverage from a substantial number of journals of medium and fringe productivity will be allocated among the member services according to the regions or languages or the journals.

The necessary studies under way or to be carried out in the near future involve (a) subject definitions and classification, (b) overlap in the abstracting of primary journals, (c) international serials data system or identification, listing, and controlling periodicals of interest to access services, (d) selection policies and overlap in coverage of individual papers within journals to assist in the development of exchange efforts, (e) editorial policies and procedures, (f) indexing approaches-similarities and differences, and (g) detailed inventory of member services operating systems to identify degrees of compatibility and convertibility.

It is anticipated that these studies can be carried out during 1971-3 and elements of the input plan would be implemented by participating services in the interim as feasible. Further details of the Input plan are available from the ICSU AB headquarters in Paris.

Finally, in regard to Part II, Processing, and Part III, Output, stages, drafts of these programs are underway. The time has not come yet for ICSU AB to launch such plans which need to be carefully studied and have the full support and back-up of its member services before being published. When these are ready to launch, they will be made as well known all over the world as had the «input plan», and comments from other organizations, components of the world system for information transfer, will be as welcome as they were for this Input plan. In developing these plans and systems for access services, the ICSU AB would be pleased to cooperate with any other organizations or individuals competent to contribute to the work.

6. Conclusion

Access services may have, and will have, a brilliant future. The future development and improvement of access services will save time, money, avoid duplication of efforts and help human progress faster along the road of knowledge for the social and economic benefit of all mankind.

Cooperation at all levels is a big venture. Two basic principles must not de forgotten when one tries to develop international cooperative activities. No plan is going to succeed, however attractive it may be, if each component of the network is not going to benefit, in one way or the other. And, secondly, a plan which would usually take two or more years to be implemented within a single organization or nation, will take at least twice as long to be implemented if it involves international cooperative activities.

Nevertheless, the opportunity for a coordinated international network of access services for the transfer of scientific and technical information increases each year. Major and current motivating forces are:

a. Access services are eager to improve their systems and services.

b. The financial pressures are forcing access services to consider ways of sharing rather than duplicating materials and resources.

c. The advances in technology of communications permit sound, pictures and digital data stored at distant locations to be made available rapidly and efficiently with relative ease.

Access services strongly believe in cooperative efforts. We hope to have convinced the readers of this paper.

A. I. Chernvi

(USSR)

INTEGRATED INFORMATION SYSTEMS

1. Definition: general aspects

A major prospect for meeting the present critical situation in the domain of scientific communications — which is a natural corollary of the advancement of scientific and technical revolution is the development and practical implementation of a special type of information systems, known as **Integrated Information Systems.** What is an integrated information system?

By an integrated information system we mean a multi-function automated information system which, once an information item has been fed into it, will repeatedly use that item in accomplishing all kinds of information tasks; it is intended for collecting, analytical-synthetic processing, storage, retrieval, and dissemination of scientific and technical information to fully satisfy comprehensive meeting of the information needs of scientists and practical specialists. In other words, exhaustive data on the contents and form of every scientific document are entered into the automated information system only once to be stored in it; this **complete** (exhaustive) **description** of the document is used by such a system as a basis for producing **all** printed aids and conducting **all** types of information service that do not imply any additional creative processing of that input information on all kinds of request.

An integrated information system is a **complex whole** made up of several simple systems, each designed for some basic function and capable of independent operation, but taken in an organic unity with the other simple systems-being an inherent part of the integral complex which can fulfil certain new functions infeasible for any of its constituent simple systems. That is to say, the total set of the functions performed by a correctly designed integrated information system must always be greater than the arithmetic total of the functions that its constituent simple systems can perform. An example of such new function of an integrated information system is its capacity for meeting practically **all** types of information needs, whereas simple systems are usually oriented at a particular type of need.

It should be emphasized that an integrated information system is **one** system, that its subsystems and components must be fully integrated, that is organically tied in and merged together. These subsystems and components make up a single whole, indivisible without a loss of some specific property.

A complete documentary integrated information system must comprise at least the following functional subsystems:

— selective dissemination of information,
— current-awareness bulletins,
— abstract journals with indexes, and
— retrospective search of documents and information.

We believe that the proper approach to building an integrated information system is not «inductive» — i. e. by identifying the common elements of the simple systems being integrated which make it possible to join or tie up these systems into a single complex—but we think that this approach should be «deductive» — i. e. by decomposing the information handled by each of the simple systems being integrated into structural elements, identifying the non-repeating structural elements, and devising a new bigger system handling only the selected minimum of structural elements but capable of all the functions performed by the initial simple systems. Thus we obtain not a «mixture» of the simple systems, but so to speak, their «solution», which is characterized by less entropy than that of the arithmetic total of those systems.

An integrated information system in principle resembles Meccano — a mechanical construction game for children, which is a set of standard parts of a few patterns and a manual showing the child how to put together various gears using the same parts.

The concept of an organic union, i. e. integration, of information systems deserves a special discussion. **Integration** is a totality of events or actions leading to the end of existence of separate parts as individual free entities and to the emergence of a previously non-existant new entity possessing certain new unprecedented peculiar properties. The notion of integration is intrinsically connected with the notion of **integrity,** for without it integration would have been synonymous to simple summation producing no qualitative change, that is the very notion of integration would have been senseless.

It is not always possible to combine separate parts into an integral whole. For that, these separate parts must possess a certain sum of properties affording the emergence of a system of connections between these parts. In relation to biological systems Acad. V. A. Engelhardt refers to such properties as **desmogenic** (derived from Greek desmos — connection') [20, p. 807].

A system of connections between the parts is a **sine qua non** of an integral whole. It is these connections which constitute the new entity distinguishing the integral whole from the total of its component parts. «Knowing everything of what 'one' is and that one and one make two, we still do not know everything about what 'two' is, for some AND is added here and it is to be known what that AND purports», as L. von Bertalanffy writes in a half-joking way [20, p. 803].

As a system of connections emerges which joins together the former free parts into an integral whole, these parts to an extent lose some of their properties; the latter are converted into the properties of the complex whole that are not identical with the lost properties of the constituent parts.

Finally, the parts constituting the integral whole — as contrasted to a mixture—must be arranged in it in a fixed and regular order.

To recapitulate, the characteristic features of an integral whole as regards its interreations to the constituent parts are:

— emergence of a system of connections between the parts constituting the integral whole;

— the disapperance of some of the part's properties as the result of its incorporation in the whole;

— the emergence of new properties in the integral whole, differing from the properties of its constituent parts; and

— the regular arrangement of the parts in the integral whole, the determination of their spatial and functional relationships.

These features have been ascertained by Acad. V. A. Engelhardt, the father of a science-cognitive trend in the present-day biology which he baptized **integratism** [20, pp. 807—808; 21, p. 10—11]. The essential ideas of that research trend have a particular value for those concerned with the theory of integrated information systems.

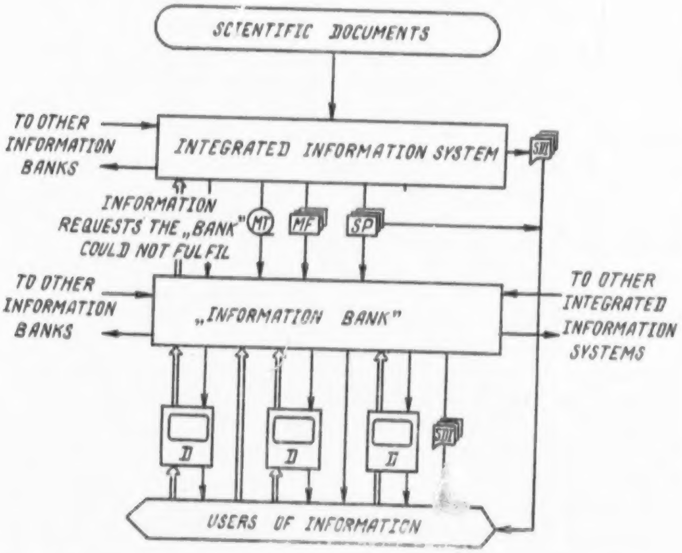

SCIENTIFIC DOCUMENTS
TO OTHER INFORMATION BANKS
INTEGRATED INFORMATION SYSTEM
SDI
INFORMATION REQUESTS THE „BANK" COULD NOT FULFIL
MT
MF
SP
TO OTHER INFORMATION BANKS
„INFORMATION BANK"
TO OTHER INTEGRATED INFORMATION SYSTEMS
D
D
D
SDI
USERS OF INFORMATION

Fig. 1 General scheme of the interaction of an integrated information system with the «information bank» and the users [double arrows show the information requests)

Abbreviations used

D—CRT display with a keyboard
MF—imicrofiches of original scientific documents
MT—magnetic tapes with indexed search files
SDI—selective dissemination of information
SP—secondary publications (current-awareness bulletins, abstract journals, etc.)

An integrated information system should be distinguished from what is called an «information bank» (or «data bank»). An «information bank» here implies a computer—based information system whose main function is to **cater to the information needs** of scientists and practical specialists **using ready-made research files** supplied to that system by one or more integrated information systems. As a general rule, the «information bank» itself performs no analytical or synthetic processing of the documentary sources of scientific and technical information. The overall scheme of interaction of a major integrated information system with a network of «information banks» connected to it is shown on Fig. 1.

2. Multidisciplinary information needs of the contemporary scientists and practical specialists

Integrated information systems have been called into existence by a social need stemming from certain peculiar features of the organization and development of modern science and technology.

An important feature of the present-day science and technology is the fact that their organization is based on at least three dimensions:

— science and technology retain the traditional divisions into disciplines:

— science and technology are organizationally arranged by the economy sectors and industries, which are not identical to the scientific disciplines; and

— science and technology are becoming more and more oriented on multidisciplinary problems, necessitating the use of the methods and results of a great number of scientific and technical disciplines.

This means that to meet the information needs of the contemporary scientists and practitians one must be able to produce any disciplinarity pattern of the information aids and services, i. s. to group or «package» the sources of scientific information by any given combination of characteristics. This task is impracticable with the conventional procedures of preparing secondary publications.

Another major feature of the present-day science and technology is the fast growth of its mass scale. The past decade alone witnessed, a 2.7-fold increase of the number of scientists in the USSR, which amounted to 985, 200 late in 1971.

Besides, by the same date, the number of diplomed engineers employed in the national economy was 2,650,000, while special ists with higher and secondary special education totalled close on 10,600,000 [13]. In this context any plans for providing that huge mass of scientists and practicians by systems of selective dissemination of information should be regarded as a vain dream.

For a long time ahead, secondary publications of various types and purposes are to remain the principal means of individual information supplies to scientists and practical specialists. This conclusion is corrobortaed by the results of a questionnaire survey of 4,000 Soviet scientists and practical specialists, conducted by «**Literaturnaya gazeta**» in 1971. These results, computer-processed at VINITI, suggest that approximately 45% of the respondents replied to the question «Through what kind of channels do you receive information valuable for you?» by the statement. «Through abstract journals, express information bulletins and other published aids of VINITI» [19]. But ever more rigid requirements are being laid to the secondary publications as regards their subject composition, coverage and preparation time.

To be sure, every scientist or specialist wants information meeting his **particular** needs moulded by his personal characteristics (qualifications, mental attitudes, etc.) and the specifics of the problem he is concerned with. This suggests that every group of scientists or specialists preoccupied with a certain scientific or technical problem should be provided with a special secondary publications. The conventional preparation procedures of a new secondary publication presuppose a new processing (indexing and/or abstracting) of the entire colossal flow of published and unpublished documents describing the recent developments in science and technology. Moreover, each time many laborous operations have to be repeated all over again in re-typing, composing, proofreading, etc. In other words, each new «packaging» of scientific and technical information tailored to new requests calls for a tremendous expenditure of skilled labour, time, and material resources. Integrated information systems seem to be the sole remedy for that wastage.

To sum up, the pressures for a further centralization of indexing and abstracting of scientific and technical literature are becoming ever more urgent today. The main factors of that are:

1. A rapid growth of the number of documents produced in any field of science or technology.
2. The increase in scattering of publications in a discipline, problem or subject field over thousands of periodicals published in different languages (a result of the growing differentiation and integration of science).
3. The growing number of interdisciplinary and multidisciplinary problems and subject areas for which secondary publications have to be provided.
4. The ever more acute need for a greater exhaustivity of coverage and reduction of the time-lags in preparation of the secondary publications.
5. The need for computer applications in quick comprehensive scientific and technical document processing and the execution of diverse information tasks. These applications are economically efficient only provided the computer is used to handle very large document flows.

3. Material premises to the building of integrated information systems

Mankind's need for integrated information systems is sharpening, but to create or implement anything—the integrated information systems in the case at hand — the community must possess adequate material capabilities. As a matter of fact, the community usually never feels a real practical need for what it is technically — at least potentially — incapable of creating or implementing. As Marx wrote, «mankind always sets itself only such tasks as it can solve; since, looking at the matter more closely, it will always be found that the task itself arises only when the material conditions for its solution already exist or are at least in the process of formation» [1].

The necessary conditions for the creation of an integrated information system are

— that the information to be handled should afford this, and

— that the adequate technical facilities should exist.

The analysis of all kinds of information service shows that they are all based on — taken in different combinations — largely the same information items — author names, document titles, imprints, classification numbers, keywords or descirptors, and the like, (Cf. Table 1). Hence, by entering into a big information system all the non-repeating information items the system needs to perform its particular functions and keeping these items within the system for a pre—specified time span, one enables such a system to an efficacious performance of the functions of a set of simple information systems, per units costs of each function in that case are lower than with a simple (specialized) system. This is ensured by sparing the multiply repeated costly input of the same information items into different specialized systems and storate of those items in these systems.

Table 1

Degree of Multiple Use of the Same Information [Information Items] for Various Types of Information Service

Information item	Secondary publication									Information service	
			indexes								
	currentawareness bulletins	abstract journal	author [short]	author [complete]	permuted	keywords	subject	clased	patent	SDI	Retrospective search of documents
1. Document type	+	+	+	+	+	+	+	+	+	+	+
2. Author name (s)	+	+	+	+	+	—	—	—	—	+	+
3. Title	+	+	—	+	+	—	—	—	—	+	+
4. Imprint	+	+	—	—	—	—	—	—	—	+	+
5. Subject heading numbers	+	+	—	—	—	—	—	+	—	+	+
6. UDC numbers	+	+	—	—	—	—	—	+	—	+	+
7. Keywords or descriptors	—	—	—	—	—	+	+	—	—	+	+
8. Subject headings	—	—	—	—	—	—	+	—	—	+	+
9. Abstract	—	+	—	—	—	—	—	—	—	+	+

The technical facilities necessary for developing an integrated information system must enable:

— the storage of great amounts of information in a form permitting further machine handling;

— automated selection, from the file stored, of all the information items possessing a given characteristic or combination of characteristics;

— the necessary arrangement of the retrieved (selected) information; and

— rapid output of the relevant information in a format convenient to the user.

All that became possible after the high-speed large-memory digital computer came on the scene.

4. Computerization: an urgent need in preparing secondary publications

Abstract Journals (AJ) with a system of indexes have been the principal medium of information supplies to scientists and practical specialists for the past 140 years. The first scientific AJ in the world seems to have been the «Bulletin universel des sciences et de l'industrie», started in Paris around the 1820s. The new service was warmly received by the learned readers, who soon began to refer to the new periodical simply as «Ferrusac's «Bulletin» in honour of its editor, who had formidable name Étienne Just Pascal Joseph François d'Audebard, Baron de Ferrussac. The baron had been an army officer, but he had resigned to devote himself entirely to science, and had conceived the ambitious plan of creating a magazine which in brief reviews would give the content of all new books and articles concerning the natural sciences, physics and mathematics. He had succeeded in engaging some eminent scientists as his collaborators, which ensured the high scientific standards of his AJ [9, p. 140].

The first specialized disciplinary AJ appeared in the field of chemistry. It was the «Pharmaceutisches Central—Blatt», initiated in Germany in 1830 and from 1856 till its closure in 1969 published under the title of «Chemisches Zentralblatt». A thousand and a half AJ in the various fields, problems and subject areas are published in the world at present.

The purely informational functions, however, are but a part of tasks fulfilled by AJ in the present-day science. AJ is part and parcel of scientific and technical literature and hence of the social mechanism of science. The main functions of AJ today are:

a) it is a medium of signalizing to scientists and specialists all scientific and technical publications in a given field or fields (current-awareness function);

b) it is a tool of retrospective search for scientific documents in the various disciplines, problem and subject areas;

c) it serves as a compensating means to surmount the publications scattering — this negative reverse side of science differentiation;

d) it promotes a considerable lowering of the language barriers;

e) it promotes the integration of science by enabling a scientist or specialist to keep abreast of the accomplishments in the allied fields of science or engineering and to make use of these advances in his own field; and in due time to identify the «juncture» fields with their active «growth points» and to switch over to these fields, etc.;

f) it helps to maintain the unity of science (through unification of special terminologies; suppression of the growth of the language barriers; elaboration of pragmatic classifications or subject schedules of sciences, etc.);

g) it serves as a means of indirect evaluation of the scientific standards of publications (the publications of a lesser scientific informativity are not covered in AJ of merely signalized by bibliographic references).

Only a half of the above functions of AJ can be viewed as purely informational, while the remaining functions are crucial to the advancement of science as such. Hence, the support and further improvement of AJ is a universal scientific commitment.

Supplemented by a system of current-awareness bulletins, AJ, as mentioned, are the major medium of information supply to scientists and specialists. However, the users today are no longer satisfied either with the current time terms of the preparation of current-awareness bulletins and AJ, or with their subject scope patterns and schedules. What is to be done to cut down the time spent on preparation of the secondary aids and to issue these aids in any discipline, problem field or subject area?

An inherent flaw of the conventional procedures still widely practised in AJ preparation is the fact that, up till the transfer of the material to the printers, the process unit being handled is a single publication (document). In the printing phase, the process unit is an array of secondary documents — abstracts, annotations and bibliographic descriptions, which together make up the manuscript of the current AJ issue. Prior to handing over the AJ manuscript to the printshop to the composed and printed, the abstracts, annotations and bibliographic descriptions are arranged (ordered) by the subject schedule adopted in the AJ series in question, and serially numbered. The serial number serve to identify the documents covered in the AJ and are necessary for the indexes.

Al publications prepared to be included in the current issue of AJ lie by after they have been edited until the day of the editing-and-publishing cycle on which the entire typescript of the issue has been fitted together to be handed over to the printshop. If, however, one could pass on to the compositor and hence to the proof-reader each abstract as soon as it has been edited, the lie-by time would be cut down to a half-day-instead of 7 or 14 calendar days, respectively, for biweekly and monthly AJ. Why then has this practice not been introduced?

The chief stumbling block is this. The large flows of documents being handled (in VINITI about 4,000 per working day) make it very difficult to make up the AJ issue and number the items in it when using conventional typesetting facilities: in fact it would take so much time and effort to ensure that the separate abstracts be not lost during the typesetting and proofreading and later during the arrangement of these abstracts according to the subject headings, to lay-out the issue and assign serial numbers to the abstracts, and to transfer these numbers to the card files of the indexes, that the procedure would be practically not worthwhile.

The only possible way to introduce a system of typesetting and proofreading of each abstract rightaway after its editing is to use a computer featuring a memory of an extremely high capacity. A batch of abstracts edited in the previous day will be recorded every day on a machine-rea-
be tantamount to the typesetting of that batch of abstracts.
dable medium-punched tape or magnetic tape, which will
Every abstract must be assigned with one or more numbers of the subject headings under which that abstract comes. The proofs are read either immediately after the recording of the abstracts on the machine-readable medium,

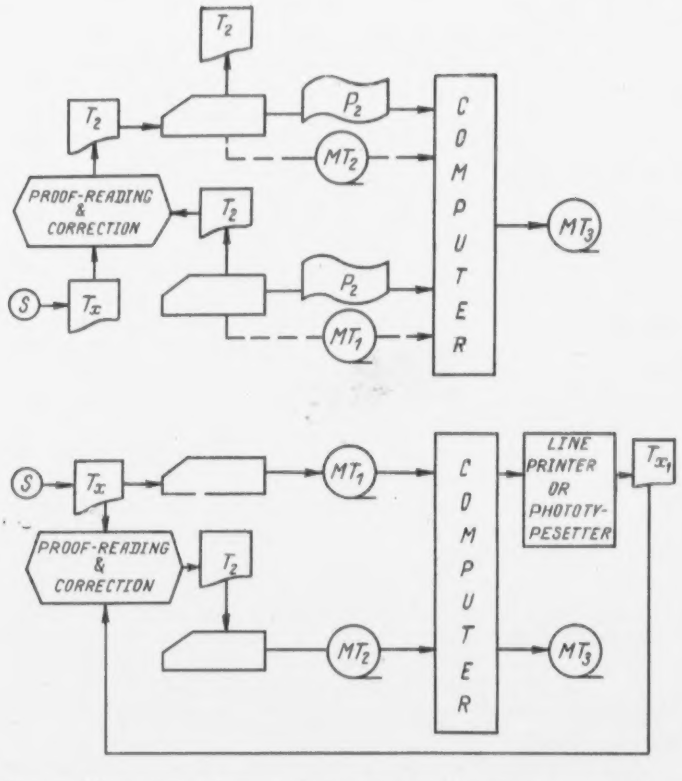

T2
T2
P2
MT2
PROOF-READING & CORRECTION
T2
COMPUTER
MT3
S
Tx
P2
MT1
S
Tx
MT1
COMPUTER
LINE PRINTER OR PHOTOTYPESETTER
Tx1
PROOF-READING & CORRECTION
T2
MT2
MT3

Fig. 2 Alternative schemes of document-by-document computer input of texts and proof reading

Abbreviations used

MT_1—magnetic tape with Tx and input errors
MT_2—magnetic tape with corrections of errors
MT_3—magnetic tape with corrected Tx
P_1—punched tape with Tx and input errors
P_2—punched tape with corrections of errors
S—start
T_1—typescript with Tx and input errors
T_2—typescript with corrections of errors
Tx—text of edited bibliographic entry, abstract, etc.
Tx_1—print-out or photo galley of Tx and input errors

or on computer-produced galleys of AJ. The possible schematic procedures of document-by-document entry of texts into the computer and of their proofreading are shown on Fig. 2.

Computer input of abstracts on a machine-readable medium should be done not on a daily but on a weekly or biweekly basis, so that parallel to the input operation one could perform selective dissemination of information. In this way ready abstracts will be stored in the computer to be included in the next issue of AJ.

When the day comes to make up a current issue of AJ, the computer automatically arrays the abstracts in its memory into the necessary sequence. For that purpose the subject heading numbers are used, which have been assigned to every abstract during the editing. Then the computer performs the serial numbering of the abstracts and prints out the AJ issue in the form of galleys. A high-speed phototypesetter (PTS) connected to the computer can be efficiently used for that purpose, since only that kind of equipment features the vast set of diverse symbols (up to 1,500) necessary for the composition of AJ in all fields of science and technology. PTS produces photo galleys, which are read by AJ editors. Through the computer, the necessary corrections can be introduced, whereafter PTS produces the final photo galleys to be used for the make-up and photo offset reproduction.

Moreover, application of computers and phototypesetters for preparation of AJ dummy is justified more than just by the essential reduction of the time of AJ preparation. Even more important is the fact that after the galleys have been printed out, **all the information entered into the computer remains there**. It is inefficient, however, to store complete texts of abstracts in binary form, for they take a tremendous storage capacity — an annual total of 10^{10} bits for VINITI. Therefore, abstracts and annotations are printed out from the computer and automatically transferred on to roll film and/or microfiches. What remains in the computer memory is complete bibliographic descriptions of the publications with their subject heading numbers. This information alone permits automatic production, by the computer-phototypesetter system, of the following information services:

— biweekly current-awareness bulletins, based on a classified systematic arrangement of material;
— author indexes; and
— permuted indexes to AJ issues.

With the conventional procedures one has to repeat the preparation of the typescript for each of the above services, as well as the composition and proofreading operations, notwithstanding that all this has been or will be done for AJ.

Moreover, by adding to the text of the abstract, in the course of its writing and/or editing, certain information elements over and above the subject heading numbers (e. g. keywords indexing the publication) one will be able automatically to prepare, on the basis of information once fed into the computer, not just AJ but also other secondary publications, and also to conduct a variety of information services, including:

— selective dissemination of information,
— preparation of current-awareness bulletins,
— preparation of subject and author indexes (to each current — awareness bulletin issue, AJ issue, as well as semiannual, annual and five — yearly).
— conduct retrospective search of documents, etc.

Besides, the approach described above brings an easy solution to the important and difficult problem of preparing secondary publications not only for the individual disciplines and fields of science and technology, but also for the interdisciplinary and multidisciplinary fields. Owing to the accumulation of the texts of abstracts, annotations and bibliographic descriptions in the computer memory, it suffices to produce a current — awareness bulletin issue or AJ on any interdisciplinary problem to make the computer, controlled by the corresponding program, to glean out of the cumutative file the pertinent texts, arrange these in the appropriate order and reproduce these as photo galleys with the aid of PTS. Hence the costs of preparation of a secondary publication on some interdisciplinary problem will be equal to the cost of the computer time spent on making up the text of the issue and on its reproduction in the form of photo galleys, plus the costs of the printing of the run. No additional investment of intelectual efforts is required.

5. Design of integrated information systems: general principles

We thus come to the concept of an **integrated information system** (IIS), the general block — diagram of which is given on Fig. 3. IIS should be designed on the basis of the principle of **one-time exhaustive analytical-synthetic processing** of each document — journal paper, progress report, patent or inventor's certificate specification, etc., — to be done by highly skilled specialists, and followed by **one-time input** of the results of that processing into the computer, for the subsequent **repeated and multiple utilization** of these results — for selective dissemination of information, preparation of required secondary publications, as well as retrospective search of documents and information. In other words, IIS «packages» information materials according to the types of information needs. The word «one-time» does not imply that at the IIS input each document is completely processed by one specialist. It only means that the processing fully excludes any duplication of intellectual efforts: it is permissible to enter into the computer only such information items that have not been previously entered into it and cannot be directly obtained or derived from those already stored.

To ensure an exhaustive analytical-synthetic processing of every document to be entered into IIS, special forms are used, which are filled in by the specialists abstracting and indexing the document. The forms are then edited and become the input documents of IIS.

A form used in IIS must include at least the following information items:

— document working number;

— document type (journal paper, monograph, patent specification, etc.);

— Russian translation of the title (for foreign-language documents);

— complete bibliographic description of the document (author's name or names, original title, imprint);

— classification numbers (by the subject heading schedule);

— list of descriptors and/or keywords indexing the document;

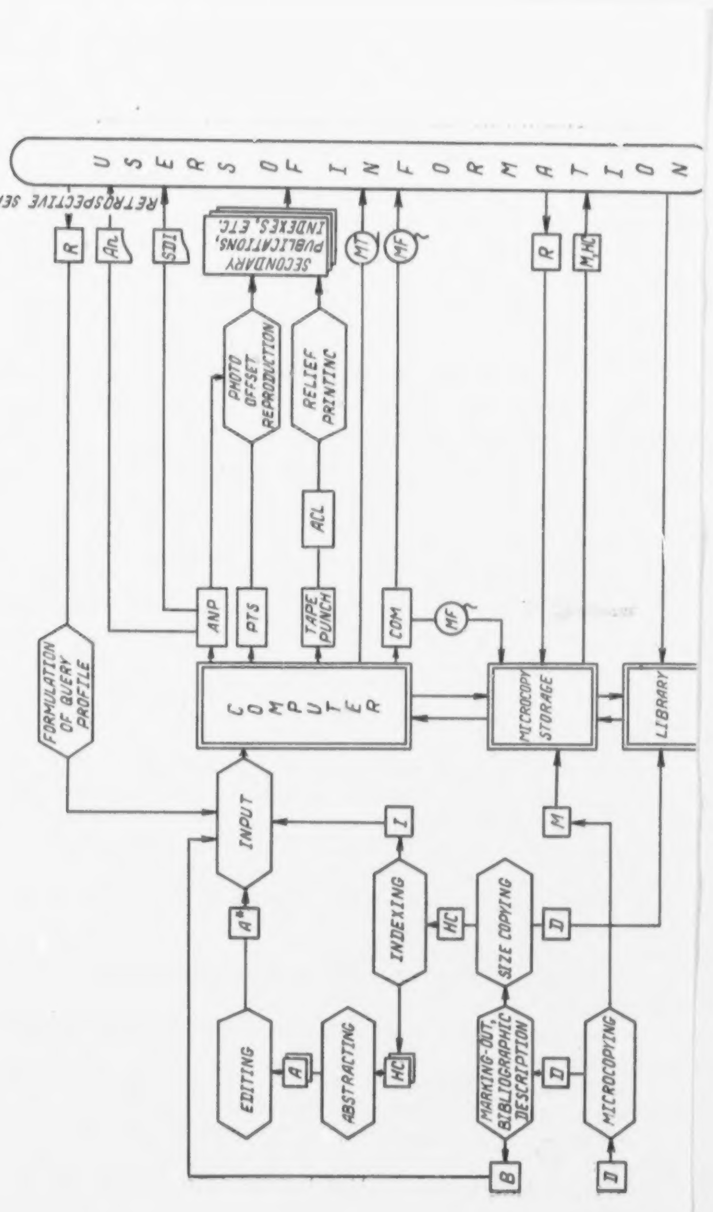

EDITING
A
ABSTRACTING
HC
MARKING-OUT, BIBLIOGRAPHIC DESCRIPTION
D
MICROCOPYING
D
B
A*
INDEXING
HC
SIZE COPYING
D
INPUT
I
M
FORMULATION OF QUERY PROFILE
COMPUTER
MICROCOPY STORAGE
LIBRARY
ANP
PTS
TAPE PUNCH
COM
MF
ACL
PHOTO OFFSET REPRODUCTION
RELIEF PRINTING
R
An
SDI
SECONDARY PUBLICATIONS, INDEXES, ETC.
MT
MF
R
M,HC
RETROSPECTIVE SEARCH
USERS OF INFORMATION

Fig. 3 Schematic block-diagram of integrated information system

Abbreviations used

A—form with document abstract
A*—form with edited document abstract
An—answer
B—form with bibliographic description and subject heading numbers of the document
ACL—automatically controlled linotype
ANP—alphanumeric line printer
COM—computer output microfilm unit
D—input periodical, patent specification, etc
HC—hard copy of document
I—form with subject heading numbers of a document marked-out, and keywords indexing that document
M—document microcopy
MF—roll microfilm
MT—magnetic tape
R—request
SDI—selective dissemination of information

— abstract.

This set of data seems sufficient for preparing almost any type of secondary publication and for meeting most types of information needs of scientists and practical specialists.

IIS embodies an essentially new approach to the design of automated information systems, striving towards the greatest possible saving of intellectual efforts spent on scientific document processing via complete elimination of all duplicate operations. That goal is achieved by enabling a multiple re-use — in various combinations — of the same information items, once fed into the computer and permanently stored there, for the automated preparation of all kinds of secondary publications and for the various types of information services.

The development of IIS pre-supposes a far better organization of work in sci-tech information agencies — highly qualified and skilled scientists and experts should be concentrated at the system input to do all the intellectual work of abstracting and indexing scientific documents.

At present IIS of this type are under development in the following information centres and institutions of the world: in the USSR—VINITI, for all the basic fields of exact, natural and technical sciences, as well as in a number of central branch institutes of scientific and technical information; in other countries — Chemical Abstracts Service (USA), for chemistry, including chemical engineering, chemical physics and biological chemistry; Institute for Scientific Information (USA), for chemistry and chemical engineering; American Institute of Physics, for physics; Institute of Electrical Engineers (UK), for physics, electrotechnology and control; Excerpta Medica Foundation (the Netherlands), for medicine; Centre de documentation du Centre national de la recherche scientifique (France), for all the major fields of science and technology; the Internationale Dokumentationsgesellschaft für Chemie (a corporation founded in 1967 by a group of chemical and pharamaceutic firms in FRG, Austria, and the Netherlands, and at present incorporating 12 companies), for organic chemistry; Japan Information Centre for Science and Technology, for all major fields of exact, natural and technical sciences; International Atomic Energy Agency (Wien), for nuclear science including marginal areas; Euratom Centre for Information and Documentation, for reactor technology, neutron physics and marginal and related sectors (such as nuclear biology and medicine, the chemistry and metallurgy of nuclear materials, applications of radioisotopes, etc.). A similar system designed for the production of «Index Medicus» (a bibliographic index in medicine, both clinical and non — clinical, and biological medicine) is operational since 1964 at the US National Library of Medicine.

6.IIS—VINITI: main features

Practical work on development of VINITI's IIS was started in 1969. This IIS has no analogue in the world; its specific features are:

a) IIS—VINITI is devised as the central system of the country for centralized analytical-synthetic processing of world's scientific and technical literature with exhaustive coverage;

b) IIS—VINITI is a multidisciplinary system encompassing all exact, natural and technical sciences and fields;

c) the development of IIS—VINITI is to be conducted **in parallel** with the further unfolding and improvement of all the existing activities of VINITI in the present organizational forms.

These specific features give rise to peculiar difficulties and problems as yet untackled in the world experience of design and practical implementation of such information systems.

The designers of IIS—VINITI must provide for:

— compatibility of this system with the other similar systems both in the USSR and abroad;

— capacity for processing 2.5 to 3.0 million publications a year, i. e. the tenfold of the biggest known IIS in the world;

— surmounting of the difficulties springing from publications scattering as well as differentiation and integration of science; and

— preservation of the existing organizational forms of work paralleled by a gradual re-training of personnel in the period of IIS implementation.

7. IIS—VINITI: basic functions

IIS—VINITI must perform the following basic functions:

a) Indexing, annotating and abstracting all world scientific and technical literature, and on that basis,—

— production of current-awareness bulletins in the various fields of science and technology, subject and problem areas;

— production of AJ «Referativnyi Zhurnal» with indexes in the various fields of science and technology, subject and problem areas;

— dissemination, on a subscription basis, of magnetic tapes with indexed retrieval files in the various fields of science and technology, subject and problem areas;

— selective dissemination of information on sci-tech literature of the world serving a certain group of patrons;

— retrospective search of documents and information on individual requests in the whole accumulated file.

b) Preparation of surveys of the type «Advances in Science and Technology»;

c) Production — on users' requests — of copies of any scientific documents reflected in current-awareness bulletins and AJ.

d) Translation — on users' requests — of foreign scientific and technical publications into Russian.

VINITI is to cease the publication of its express — information bulletins («Ekspress Informatsiya»), transferring that function to the central branch institutes of scientific and technical information.

The current-awareness bulletins will cover all documents relevant to the field, subject or problem area concerned, without any exception. The bibliographic descriptions in current-awareness bulletins will be arranged in a systematic order. Each current-awareness bulletin issue will be furnished with an author index and a list of source periodicals.

AJ should include only abstracts and annotations but not single bibliographic description. Any duplication of abstracts and annotations between the different series and issues of AJ is admissible. Each AJ issue must be furnished with an author index and an alphabetic subject index. Cumulative indexes should be semi-annual, annual and triennial.

IIS—VINITI is to perform its basic functions under the following time terms:

a) Publication of biweekly current-awareness bulletins — with a time-lag of 20 to 25 days after the accession of the source publications to VINITI (including the two weeks during which materials are accumulated for the current bulletin issue).

b) Coverage of at least 50% of all incoming publications in AJ — after 50 days.

c) Production of copies of all sci-tech documents on users' requests — in 10 days.

8. The functional structure of IIS—VINITI

In order to cope efficaciously with the above functions observing the required time terms, IIS—VINITI must comprise 5 specialized subsystems (Fig. 4):

1. Literature processing subsystem, responsible for the indexing and abstracting of world's scientific and technical literature, and performing, on that basis,—

— preparation of current-awareness bulletins,

— preparation of AJ with indexes,

— selective dissemination of information,

— prepartaion of surveys of the type «Advances in Science and Technology»;

— on-subscription dissemination of indexed search files (on magnetic tape) in any field of science and technology, subject or problem area.

Most of these functions are being at present carried out by VINITI Institute.

2. Printing subsystem, responsible for the publishing and distribution of information materials being prepared in the Literature processing subsystem. This functions will be performed — as it is at present — by VINITI's Printing Plant.

3. Information service subsystem (or «information bank»), which will conduct retrospective search of documents and information on individual requests. At present there is no such system in VINITI, the only exception being the information retrieval system is information on science having a file of 35,000 documents.

4. Photocopy service subsystem, producing copies of publications on users' requests. This function is now fulfilled by the so-called Reference Information Centre of VINITI's Printing Plant (RIC).

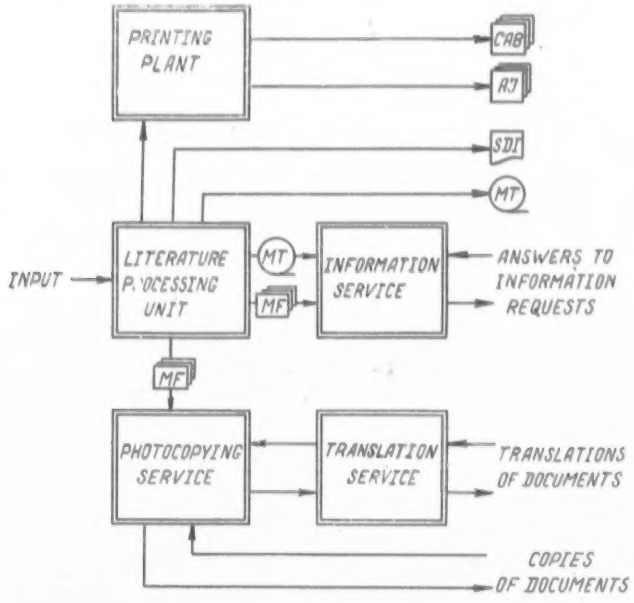


Fig. 4. Aggregated functional structure of VINITI's integrated information system

Abbreviativen used

AJ—abstract journals
AB—current—awareness bulletins
MF—microfiches of original scientfic documents
MT—magnetic tapes with indexed search files
SDI—selective dissemination of information

5. Translations service subsystem, which will produce Russian translations of foreign scientific and technical documents on users' requests. This function will continue to be performed by the All-Union Centre for Translation of Scientific and Technical Literature and Documentation (till May 1972, called «Translations Bureau of VINITI»).

It should be emphasized the Information service subsystem («information bank») sould specialize on service of the users. The retrieval files will enter that system in a form ready for use from the Literature Processing subsystem. This will enable the staff running the Information service subsystem to concentrate on the analysis and quick meeting of the information requests of scientists and specialists, and on the constant improvement of the efficiency not only of that subsystem but also of the Literature Processing subsystem.

Photocopy service subsystem — in contrast to RIG of VINITI's Printing Plant — should be an organizationally independent and nonprofit agency with the sole function of quick fulfilment of users' orders for copies of scientific and technical publications. For that, the Photocopy service subsystem must have a complete file of all documents covered in VINITI's informational publications. This file must be on microfiches of 105×148 size (reduction ratio — 24 : 1/; these will come to the system from the Literature Processing subsystem.

9. Information-retrieval languages of IIS—VINITI

IIS—VINITI will make use of three information retrieval languages:

— the system of classifications that makes up the Subject Heading List;
— UDC; and
— descriptor language.

The Subject Headings List is used for automated make-up of informational publications, and also for selective dissemination of information and retrospective retrieval. The List has such a schedule and content which make it possible to issue current-awareness bulletins and AJ issues by any desired combinations of characteristics, provided this is economically justified, i. e. there is a sufficiently large group of users with similar information requests. That means that the List should be annually updated and revised to take cognizance of the findings of analysis of the performance of the system of selective dissemination of information and of the individual information requests. The List can be elaborated on the basis of the existing «Rubricator of abstracting journals of the

USSR» [10, 11], created in the past 15 years of VINITI's activities, and, apparently, providing for an adequate reflection of the information needs of a great proportion of Soviet users.

The descriptor language is intended for compilation of alphabetical subject indexes to AJ, and for selective dissemination of information and retrospective retrieval. The vocabulary of the descriptor language is constituted by a fixed list of descriptors and any keywords, selected directly from the text of documents being indexed and functioning as descriptor modifiers. Only links will be used in document profiles. The query profiles will be formulated using Boolean operators, AND, OR, NOT, and weights.

To raise the efficiency of indexing and information retrieval, a system of compatible thesauri in various scientific and technical fields, and subject or problem areas is planned.

10. IIS—VINITI: processing procedures

The designers of IIS—VINITI proceeded from the following assumptions:

a) The main function of the Literature processing subsystem is to prepare secondary publications (current-awareness bulletins, AJ with indexes, etc.) and, on that basis,— indexed retrieval file on magnetic tapes and discs. All other kinds of information service (except for the selective dissemination of information) should be conducted by the Information service subsystem using that file as well as by other specialized subsystems of IIS—VINITI.

b) IIS should perform its functions on the basis of an exhaustive utilization of the single flow of scientific and technical publications. Duplicate subscription to foreign periodicals and other sources should be reduced to a minimum. At the same time, losses of such sources during processing must be ruled out.

c) Data on every document must be entered into IIS not as a single item, but by parts. First, bibliographic descritions of the documents with subject heading numbers assigned to them must be entered. This information takes a minimal time to prepare, though it permits to start the servicing of the users long before current-awareness bulletins are issued to say nothing of AJ. Nest, search profiles of the document (lists of keywords and UDC indexes) must be entered into IIS, since the deep indexing of documents takes far more time than the preparation of their bibliographic descriptions and their indexing by the Subject Headings List. And only at the last step the texts of annotations and abstracts should be fed into IIS, for these take the longest time to produce.

d) There should be no time limit for the storage of information in IIS.

The technological process in IIS—VINITI will be approximatingly made up of the following sequence of procedures (see Fig. 5):

1. The flow of sci-tech literature passes through the conventional accessioning and registration procedure. During the registration, a special blank form is attached to each publication-journal, book etc. — giving the abbreviated title of the publications, title code, year of publication, volume and issue number.

2. Immediately after the registration (i. e. prior to marking-out), all periodicals and serials are **copied in complete text** on to 105×148 mm microfiches, which pass over to the document storage unit of the Information service subsystems, while the copies of these microfiches are passed to the Photocopy service subsystem.

3. After copying, all publications appear on the new accessions display twice a week, alphabetically arranged on racks and available for scanning and partial marking-out by the specialists from the different editorial departments of VINITI.

4. From the display, all publications go to a special department responsible for selection and complete marking-out for current-awareness bulletins and AJ of all documents included in these publications. Each document is assigned with one or more Subject Heading numbers (down to a most narrow subdivision). Besides, the marked-out documents are selected for the subsequent coverage in AJ.

5. The source publications, already marked-out, are passed to the division where they are copied 1 : 1 on to paper on Xerox-720 or similar equipment. Specially, for current-awareness bulletins a hard copy of the first page of each document marked-out (together with a copy of the

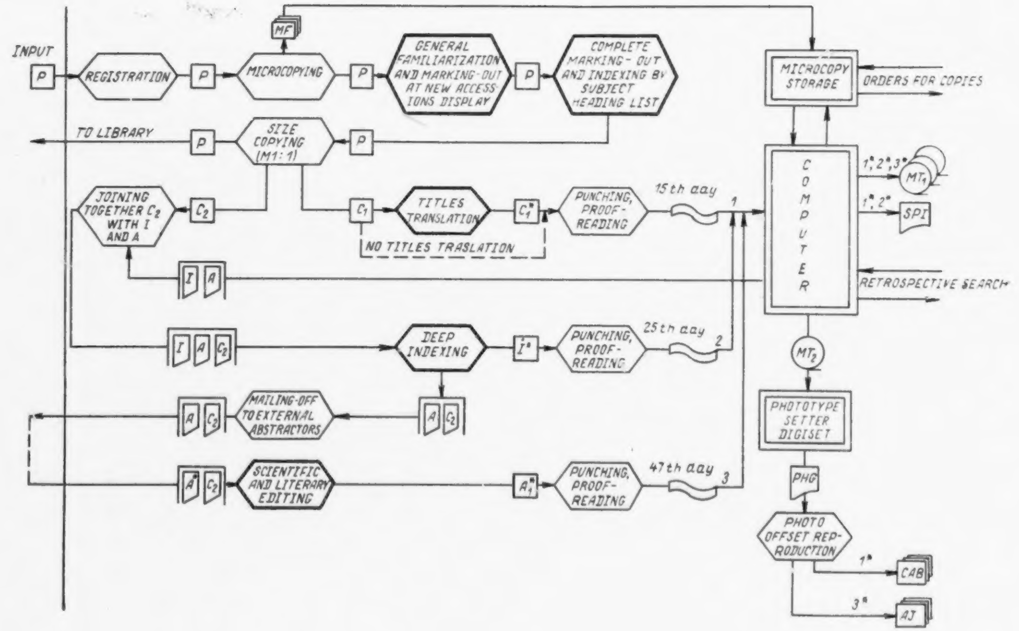

INPUT
P
REGISTRATION
P
MICROCOPYING
MF
P
GENERAL FAMILIARIZATION AND MARKING-OUT AT NEW ACCESSIONS DISPLAY
P
COMPLETE MARKING-OUT AND INDEXING BY SUBJECT HEADING LIST
MICROCOPY STORAGE
ORDERS FOR COPIES
TO LIBRARY
P
SIZE COPYING (M1:1)
P
JOINING TOGETHER C2 WITH I AND A
C2
C1
TITLES TRANSLATION
C1*
NO TITLES TRASLATION
PUNCHING, PROOF-READING
15th aay
1
COMPUTER
1*, 2*, 3*
MT1
1*, 2*
SPI
I
A
RETROSPECTIVE SEARCH
I
A
C2
DEEP INDEXING
I*
PUNCHING, PROOF-READING
25th aay
2
MT2
A
C2
MAILING-OFF TO EXTERNAL ABSTRACTORS
A
C2
PHOTOTYPE SETTER DIGISET
A*
C2
SCIENTIFIC AND LITERARY EDITING
A1*
PUNCHING, PROOF-READING
47th aay
3
PHG
PHOTO OFFSET REPRODUCTION
1*
CAB
3*
AJ

Fig. 5. Aggregated technological structure of VINITI's integrated information system. Fat lines show operations performed by specialty experts. Numbers 1*, 2* and 3* stand for types of secondary publications and/or information services that can be produced by the system rightaway after input of information 1, 2 or 3

Abbre,viations used

A—ABSTRACT FORM (WITH BIBLIOGRAPHIC ENTRY AND WORKING NUMBER OF THE DOCUMENT)
A*—ABSTRACT FORM WITH NON-EDITED ABSTRACT
A_1^*—ABSTRACT FORM WITH EDITED ABSTRACT
AJ—ABSTRACT JOURNAL
C_1—HARD COPY OF THE FIRST PAGE OF THE DOCUMENT
C_1^*—HARD CODY OF THE FIRST PAGE WITH RUSSIAN TRANSLATION OF TITLE
C_2—HARD COPY OF THE WHOLE DOCUMENT
CAB—CURRENT-AWARENESS BULLETINS
I—INDEXATION FORM NOT FILLED-IN (WITH BIBLIOGRAPHIC ENTRY AND DOCUMENT WORKING NUMBER)
I*—INDEXATION FORM FILLED-IN
MF—MICROFICHE
MT_1—MAGNETIC TAPES WITH INDEXED SEARCH FILES
MT_2—MAGNETIC TAPES WITH TEXTS OF CURRENT-AWARENESS BULLETINS AND ABSTRACT JOURNALS
P—PERIODICAL
PHG—PHOTO GALLEYS
SDI—SELECTIVE DISSEMINATION OF INFORMATION

blank form carrying the imprint data of the source publications) is produced.

6. The hard copies of the opening pages of all documents marked-out go to the division responsible for recording on punched tape or magnetic tape all information items necessary for automated production of current-awareness bulletins. Each document is assigned a working number. If all foreign titles are to be supplied with Russian translations in current-awareness bulletins, then, prior to perforating punched tape or typing onto magnetic tape, the copies of the opening pages are to be for a short time submitted to the editorial departments.

7. Information recorded on punched or magnetic tape is entered into the computer and after the control and correction procedures is stored there for the subsequent preparation of current-awareness bulletins, AJ and for other purposes. Besides, two blank forms are automatically filled in for each document selected for abstracting—**indexation form** and **abstract form.** Each form carries the working number and the complete bibliographic entry of the document plus its subject heading numbers.

8. Indexation and abstract forms together with the complete copies of the documents go to the appropriate editorial departments, where each document is UDC- and keyword-indexed. The indexation form filled-in is passed on to be recorded in machine-readable form for computer input, and the abstract form is sent out for abstracting together with the document copy.

9. The form returned by the abstractor carries an abstract of the document; it is subjected to scientific and literary editing, whereafter it is to be recorded in machine-readable form for computer input.

10. After a certain number of documents have accumulated in the computer, it automatically arrays them by the subject heading numbers, provides with serial numbers, compiles an author index and other indexes and records the resulting text on magnetic tape in a form necessary for photocomposition.

11. A DIGISET phototypesetter is used for the composition operation. After the galleys have been read and corrected, the dummy is made up. The final copy is signed for press by the editor of the current-awareness bulletin or

AJ issue, and sent to the printing plant for photo offset or electrographic reproduction.

12. The run of the informational publication is produced and distributed among the subscribers.

The major advantages of the above procedure are as follows:

a) Multiple typesetting and proofreading of the same information items is fully excluded (such as abstract numbers, author names and complete bibliographic descriptions), it is no longer necessary to keep numerous and cumbersome card files, etc. The result is a tremendous saving of labour, time and materials, and a considerable reduction of errors.

b) A capability is given for selective dissemination of information as well as a retrospective search of documents (by elements of the bibliographic descriptions and the subject heading numbers) as soon as after 15 or 20 calendar days since the accessioning of the source publications at VINITI (i. e. for, example, during the input of the file of bibliographic descriptions accumulated over two weeks for current-awareness bulletins).

On the 25th or 30th day after the accessioning of the source publication at VINITI, i. e. after the computer input of the indexation forms, that capability becomes unlimited.

c) Automatic compilation of author, alphabetic subject, patent and other indexes to each AJ issue becomes possible.

d) A practical opportunity is created for the issuance of secondary publications in any field of science of technology, and any subject or problem area in very short time terms.

11. IIS-VINITI: hardware

The implementation of IIS—VINITI should be based on a set of technical facilities, including:

— a «third-generation» computer in the appropriate configuration,

— a high-speed phototypesetter,

— a fleet of keyboard devices for the recording of input information on punched cards, punched tape and magnetic tape,

— equipment for the storage and retrieval of the microcopies of input documents («second loop» of IIS),

— a fleet of electrographic copiers of the type Xerox-720, -3600, -4000, etc.

Rough calculations show that IIS — VINITI needs a computer featuring:

— a ramified time-sharing system, accomodating up to 50 or 60 terminals, including at least 20 CRT displays, and a response time of 3 to 5 secs;

— a powerful central processing unit featuring 0.5 to 1.0 million operations per sec.;

— a memory on interchangeable magnetic discs with the total capacity of up to 600 million bytes and access time of some tens of milliseconds;

— a core storage with a capacity of up to 400,000—500,000 bytes, with an access time of 1 or 2 μsec;

— storage units on magnetic tapes (no less than 8 decks);

— the necessary number of input and output devices handling punched cards, punched tapes and/or magnetic tapes;

— alphanumeric output chain printers having a set of not less than 192 characters and a speed of 600—1000 lines/min.;

— COM devices for direct output of information on to microfiches and roll film;

— communication processor; and

— visual displays and terminal teleprinters.

The displays and teleprinters, connected to the computer, can be installed at various locations across the country, at distances of hundreds and even thousands miles from the computer. These terminals will give the user a direct access to IIS—VINITI and immediate reception of relevant information. Displays and teleprinters are particularly effective for retrieval purposes. They afford the user-system dialogue so that the request statement can be modified by the intermediate retrieval results to ensure the best end results.

To provide a great number of users with an opportunity for holding such «dialogue» with the IIS simultaneously, it must be based on a computer system with a developed timesharing facility. The user should not have to stand in a queue waiting until the computer has answered an earlier request of another user. Time-sharing will be absolutely indispensable for the economically efficient performance of IIS—VINITI and the prompt answering of requests.

12. The current status of IIS — VINITI

At present, VINITI has four computers (two Minsk-22 and two Minsk-32), which are heavily loaded with various practical tasks (in particular, preparation of current-awareness bulletins in automatics, radio-electronics and chemistry), information retrieval in the field of information science, preparation of various indexes, and elaboration and updating of abstract typesetting programs, etc. These computers have a sufficient total capacity for production of secondary publications of an overall volume equal to 45% of the totality of present, VINITI's secondary publications.

To raise the performance of these computers they will be joined together — by means of buffer magnetic core storages — to make a single complex system, which will greatly expand their technical capabilities. The merger of the computers into a single system will be completed in 1973. In 1971, a magnetic-tape SC-1060A storage was connected to the MINSK-22 computer, enabling it to read out magnetic tape sent from abroad and to record information on such magnetic tapes.

In 1972, a MD-17 storage on interchangeable magnetic discs will be connected (each pack of discs featuring a capacity of 7.25 million bytes) as well as an alphanumeric Potter NCR-3502 chain printer with a set of 192 characters. This hardware permits to work out the necessary software, accumulate the valuable technological experience, and to carry on the necessary personnel training. In this way VINITI will be able in a very short time to put into operation a large third-generation computer envisaged for the implementation of the final version of IIS—VINITI.

VINITI's DIGISET 50T-1 phototypesetter has been in regular operation since mid-February 1970. Its working speed is 300 to 600 characters/sec (depending on the size of characters). Some 800 different characters have been entered into the unit's memory, including 140 Cyrillic and 86 Roman characters (in fine and semi-bold founts). All the types have been developed at VINITI. Current-awareness bulletins in automatics and radio-electronics have been composed on the DIGISET machine for about two years. Moreover, a few subject and author indexes to some of the union volumes of AJ are already being photocomposed on that unit.

Computer input is mostly on punched tapes produced on Dura Mach-10, Supertyper, Optima-527 and -528 typewriters.

The debugging of the computer programs of compilation of current-awareness bulletins in the integrated processing mode, that is on the basis of a one-time computer input of information, is nearing its termination. The first photo galleys attest to the practical viability of these programs, already providing for a higher photocomposition quality than many of foreign IIS. Before the end of this year, the current-awareness bulletins for all fields of chemistry and chemical engineering will be switched over to photocomposition (38 series, about 230,000 entries in 1972). Since 1972, current-awareness bulletins in all fields of automatics and radio-electronics (11 series, about 50,000 entries a year) have been started.

Computerizing current-awareness bulletins is merely an opening phase in creating IIS—VINITI. Technologically it will be completed within the current year. The second phase will be the changeover to the photocomposition of AJ. The computer programs for that photocomposition will be written and debugged also in 1972. From 1973, at least one AJ series will be produced in that manner.

13. Building a nationwide network of scientific and technical information on the basis of integrated information systems

The objective of IIS—VINITI is not confined to the raising of the efficiency of VINITI's scientific information activities, important as it is.

We think that the creation of IIS—VINITI will affect the entire nationwide system of scientific and technical information and give an impetus towards the creation in the country of an entire network of fully compatible and interconnected functional and branch IIS.

Specialization of any kind of activity, including scientific information work, is a major means for raising its efficiency, as it permits to develop a more advanced technological procedure, make a wide use of up-to-date mechanization and automation facilities, and to train high-skilled personnel.

Scientific and technical information agencies are faced with two major categories of functions:

— analytical-synthetic processing of sci-tech information, and

— information services.

It appears, therefore, reasonable to specialize the all-union /functional/ and central branch information agencies on annotating, abstracting and indexing of scientific documents, charging the information servicing of scientists and other specialists on «information banks», which should receive their reference files in a form ready for use /e. g., on magnetic tapes/. That trend is already discernible in the organization pattern of the system of scientific and technical information agencies in a number of countries.

The basic units of the nationwide scientific and technical information network in the USSR are — and will be — the scientific and technical information departments (STID) in R&D and D&D institutions and the technical information bureaux (TIB) at large industrial enterprises (late in 1971, STID's and TIB's across the country totalled about 9,000); in addition, there are the interdisciplinary regional scientific and technical information agencies (sci-tech information centres in the territories, regions, autonomous republics and autonomous regions of the RSFSR, 62; in the Ukraine, 6; and in the Kazakh Republic, 4; republican institutes of scientific and technical information, 15), since these bodies are directly connected with science and production, for it is by them that scientists and engineers are served.

The regular operation of these information agencies is to be supported by the all-union (functional) institutes (10), and central branch institutes of scientific and technical information (82) /5, p. 13/. By functional here and henceforward we mean an information agency which fulfils a certain information function (one or more) in the interests of the whole nationwide system of scientific and technical information. An example of a purely functional information agency is the All-Union Scientific and Technical Information Centre (VNTITs) resposible for the collecting, processing, storage, retrieval and dissemination of information about all started, implemented and terminated R&D and D&D projects in the country. VINITI is both a functional (analytical-synthetic processing of all published sci-tech literature, methodological guidance of all research toward improvement and advancement of the national information service network and coordination of this research in the country) as well as a central branch institute of scientific and technical information (servicing the system of the USSR Academy of Sciences).

At present STID, TIB and regional information agencies are spending a great part of their efforts not on servicing scientists and practicians, but on abstracting and indexing the relevant documents, publishing their own abstract bulletins, keeping their own reference files, etc., that is on performing the functions which are charged on the all-union (functional) and central branch institutes of scientific and technical information. This is due mostly to the fact that STID, TIB and regional information bodies are often dissatisfied with the subject schedules, coverage and/or time terms of the information aids published by the all-union and central branch institutes of scientific and technical information.

If STID, TIB and regional information agencies were to be regularly supplied with current-awareness and abstracting publications having the necessary subject schedules and coverage, and produced in the required time terms — which could be achieved only **via** creating compatible IIS in all the all-union (functional) and central branch institutes of scientific and technical information— and if provisions were to be made for quickly meeting the information requests necessitating a retrospective search and quick execution of orders for photo copies of the primary documents, then STID, TIB and regional information agencies could largely concentrate on more advanced forms of analytical-synthetic processing of scientific and technical information — notably, on writing of analytical surveys, developmental forecasts, information synthesis, etc., and on a more profound and exhaustive study and meeting of the individual information needs of scientific and special users.

The optimum structure of the network of scientific and technical information agencies in the USSR can be visualized as follows. A nationwide system of functional and branch information processing centres is to be created, largely corresponding to the existing all-union and central branch institutes of scientific and technical information. These centres will be responsible for the analytical — synthetic processing of the relevant documentary information, striving to the greatest possible reduction of unjustified duplication of intellectual efforts spent in other similar centres: VINITI will process all published scientific and technical literature, VNTITs — all data on started, implemented and terminated R&D and D&D projects, central branch institutes of scientific and technical information — the materials of the «upward» flow, i. e. such unpublished documents which reflect the results of R&D and D&D work, data on new industrial articles, advanced know-how, etc. Indexed abstracts of scientific documents will be channeled from the information processing centres to the «information banks» on magnetic tape. The main function of these banks will be to conduct information servicing of scientists and practical specialists.

This approach means that every central branch institute of scientific and technical information must abstract and index only all unpublished documents from its particular «upward flow», updating the reference file it keeps with fragments of ready-made search files necessary for the branch of economy in question. These files should be supplied on magnetic tapes from VINITI (for the published sources), and from VNTITs and other central branch institutes of scientific and technical information (on unpublished sources). The main goal in that case is to bring down to a minimum the unjustified duplication of intellectual efforts spent on the costly analytical — synthetic processing of the same scientific documents, but not to reduce the necessary diversity of the secondary publications and services.

It is necessary to build up in the USSR a rational network of «information banks», taking into account the economic regions pattern and organized on an hierarchichal structural principle. The upper level in that hierarchy will be formed by the «information banks» presently under development at the all-union (functional) and central branch institutes of scientific and technical information; the medium level will be the «information banks» to be created at the central bureaux of scientific and technical information and the republican institutes of scientific and technical information (the regional bodies). The next hierarchical level will be the «information banks» created at the leading STID and TIB, whereas the lowest level of this hierarchy will be the reference information collections already existing in all STID and TIB. Although these collections — as well as the reference information collections of the regional information agencies and of the leading STID and TIB — are not yet called «information banks», they virtually exist and fulfil all the principal functions of such banks.

The hierarchy of the «information banks» is to form an automatic system of «filters», allowing to the next, higher, level only such informtaion requests as cannot be met by the lower-level «bank». This system of «filters» will be indispensable for ensuring the high effectiveness of the entire nationwide information network. In other words, the practice should be introduced of accepting the requests only at the «banks» of the lowest hierarchical level, i. e. at the appropriate STID or TIB, except for specially stipulated cases. That practice should be strictly observed. A request will then be sent to a higher-level «bank» only by the starting lower-level bank if the latter is itself incapable of meeting it. However, the response will be sent directly to the user, independent of the hierarchichal level of the respondent «bank».

Every «bank» must continually study the information needs of its users, and accordingly to update its reference file with the accessions incoming from one or more information processing centres.

The «information banks» of all the above categories virtually exist and are functioning in the USSR. The task now is to ensure the necessary compatibility of these «banks» and to set the system of their interaction going according to the principles set forth above.

It will be possible to organize a unified system of information processing centres and «information banks» only if these are equipped with compatible third-generation computers. The computers must have special peripherals, including a large external memory and CRT displays; they must operate in the time-sharing mode. Thereby the opportunity will be opened for connecting together these computers — through the usual communications channels — into one unified system, hierarchically organized. That system — thanks to teleprinter and CRT displays — will provide practically every scientist or practician with a direct access to any «information bank».

The all-union network of photocopying centres for scientific documents should have roughly the same localization pattern, structure and interaction mechanism. The 105×148 mm sized microfiche holds a great promise as an information medium and it should be broadly introduced into the practice of information work. At a reduction ratio of 20 : 1, one microfiche holds from 60 to 72 pages, and at a reduction of 60 : 1, as many as 475 pages, that is a whole book of the average size; if the reduction is 150 : 1, one microfiche holds up to 3,200 pages, that is 7 or 10 books! Such microfiches are usually kept in catalogue drawers, where one linear centimetre holds up to 40 of them.

When special photo materials which do not contain silver (e. g. diazofilm) are used, the cost of microfiches is by far lower than that of printed originals, to say nothing of the costs of their enlarged hard copies. If the users are provided with the sufficient number of inexpensive readers and reader-printers, this will bring an efficient solution to many involved problems: the furnishing of the regional centres with the complete files of the publications and documents they need, publishing certain types of services directly on such microfiches, making foreign literature by far more available to the users, finding a drastic solution to the storage space problem, etc.

Networks of «information banks» and photocopying centres should be supplemented by a nationwide information referral service, i. e. a network of agencies capable of giving the user a quick answer as to which centre he should refer to for scientific and technical information he needs. Besides, the existing network of «information analysis centres» should be substantially enlarged — these are bodies created at the leading R&D and D&D institution and intended to write critical reviews and analytical surveys, deve-

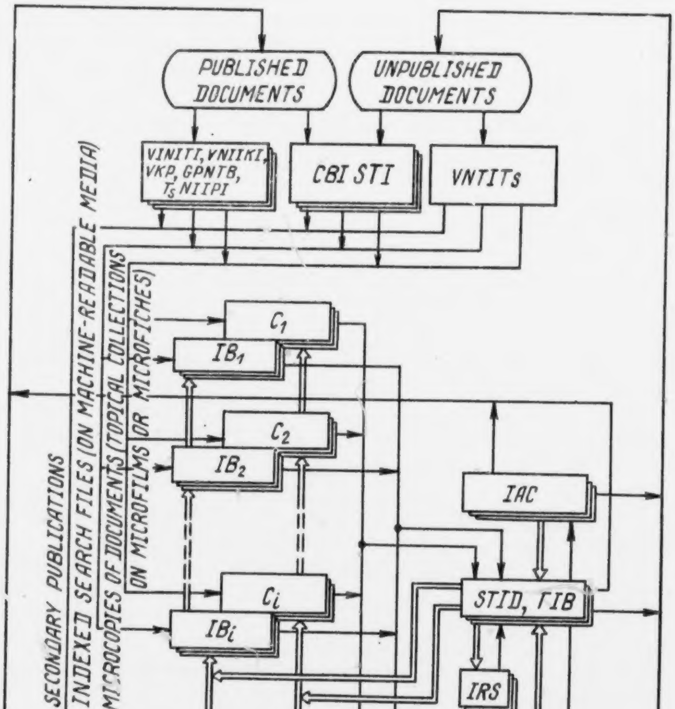

PUBLISHED DOCUMENTS
UNPUBLISHED DOCUMENTS
VINITI, VNIIKI, VKP, GPNTB, TsNIIPI
CBI STI
VNTITs
SECONDARY PUBLICATIONS
INDEXED SEARCH FILES (ON MACHINE-READABLE MEDIA)
MICROCOPIES OF DOCUMENTS (TOPICAL COLLECTIONS ON MICROFILMS OR MICROFICHES)
C1
IB1
C2
IB2
Ci
IBi
IAC
STID, TIB
IRS
SCIENTISTS AND PRACTICAL SPECIALISTS

Fig. 6 General functional scheme of the nationwide scientific and techunical information network based on integrated information systems (double arrows show information requests)

Abbreviations used

C_1, C_2, ..., C_i—photocopying centres
CBISTI—central branch institute of scientific and technical information
GPNTB—State Public Scientific and Technical Library
IAC—information analysis centre
IB_1, IB_2, ..., IB_i—«information banks»
IRC—information referral centre
STID—scientific and technical information department (at an R&D or D&D institution)
TIB—technical information bureau (at a big industrial plant)
VINITI—All-Union Institute of Scientific and Technical Information
VKP—All-Union Book Chamber
VNIIKI—All-Union Research Institute of Technical Information, Classification and Coding
VNTITs—All-Union Scientific and Technical Information Centre
TsNIIPI—Central Research Institute of Patent Information

lopmental forecasts, etc. Specialists of the highest qualifications should be enrolled to work for such centres; they must be recruited from among active researchers and designers. The overall functional scheme of the nationwide network of scientific and technical information agencies is given on Fig. 6.

Special attention should be given to the problem of information supplies to the managerial agencies of the country. The crux of this problem lies, as we believe, in the fact that while, on the one hand, the government agencies are literally flooded with an information flow they are unable to digest because of its monstrous volume and diversity, on the other, the state management still does not receive the sufficient amounts of information expressly prepared and tailored to its needs and fit for direct use in managerial decision making.

To solve that problem it will be necessary to set up a limited number of special information analysis centres subordinated, respectively, to the USSR Council of Ministers, the State Planning Committee, the State Committee for Science and Technology, and the Presidium of the USSR Academy of Sciences. Experienced scientists and experts of high renown should be enlisted to participate in the activities of these centres. These centers should be endowed with the exclusive authority for supplying information to the supreme government agencies, whereas the latter must have an opportunity for quickly obtaining any data they need from the information bodies of any level or affiliation.

To ensure a fuller utilization of available scientific and technical information in R&D and D&D projects, it seems sensible to examine the feasibility and efficiency of introducing the practice **of information «defence»** of a proposed project. The information agency concerned should be in that case made responsible — along with the proponent of the project — for the consequences of a failure to use all available scientific and technical information at the start of a project.

The conversion of science into a direct productive force of the contemporary society confronts the latter with the tasks of a considerable increase of the efficiency of research and design and development, and a reduction of the time taken to implement the results of these in practice.

In the complex chain connecting science and industry the weakest are the links responsible for the practical implementation of the results of science, for their mass-scale introduction into industrial practice. To surmount the difficulties in this field a complex set of different measures should be undertaken including the constant improvement of the performance of the nationwide system of scientific and technical information. For that, it is necessary to finish the merger of the information agencies into a system of interconnected and interacting subsystems, which is to be based on the principle of **specialization.** In such a system, the abstracting and indexing of the «downward» and «upward» flows of scientific documents will be carried out by the network of functional and branch integrated information systems. The development of these systems is more and more becoming, as we believe, one of the main lines in the advancement of the nationwide network of scientific and technical information agencies.

REFERENCES

1. K. Marx. Preface to «A contribution to the critique of political economy. In: K. Marx and F. Engels, Selected Works, v. 1. Moscow, Foreign Language Publishing House, 1962, p. 363.
2. A. D. Barsky, V. I. Danilin, M. F. Korotyaev. Principles of design of integrated systems of data processing. «Ekonomika i matematicheskie metody», 1971, 7, No. 4, 603—610 (In Russian).
3. Integrated data processing systems. Moscow, «Nauka», 1970, 113 c. (In Russian).
4. R. **Johnson, F. Cast,** D. Rosenzweig. The theory and management of systems. 2nd ed. Moscow, «Sov. radio», 1971, 647 pp. (Russ. ed.).
5. O. V. Kedrovsky. Nationwide system of scientific and technical information in the USSR. In: «Nauchno — tekhnicheskaya informatsiya. Istochniki, poisk, ispol'zovanie» (A course of TV lectures). Moscow, VINITI, 1972, pp. 12—17. (In Russian).
6. A. Mikhailov. The All-Union Information System. «Pravda», March 29, 1971, No. 88 (19231), p. 3. (In Russian).
7. A. Mikhailov. Scientific and technical information and science productivity. «Kommunist», 1971, No. 16, 98—107. (In Russain).
8. S. L. Optner. Systems analysis for business and industrial problem solving. Moscow, «Sov. radio», 1969, 215 pp. (In Russian).
9. O. Ore. Niels Henrik Abel. Mathematician extraordinary. Minneapolis, University of Minnesota Press, 1957, 277 pp.
10. Subject index to the Rubricators of abstract journals of the USSR, Moscow, VINITI, 1971, 221 pp. (In Russian).

11. Rubricator of the abstract journals of the USSR, Moscow, VINITI, 1969, v. 1, 533 p.; v. 2, 60 pp. (In Russian).
12. V. P. Rumyantsev, V. A. Skripkin, L. N. Sumarokov. Models of integration of scientific and technical information systems. «Elektrotekhnicheskaya promyshlennost». Ser. Obshcheotraslevye voprosy», 1972, No. 6, 382, 21—25. (In Russian).
13. USSR in figures, 1971. A brief statistical yearbook. Moscow, «Statistika», 1972, pp. 70, 188, 199.
14. L. N. Sumarokov, Yu. M. Gornostaev. Integrated information systems: general concepts, problems and methodological aspects of design. Moscow, Informelektro, 1972, 60 pp. (In Russian).
15. A. I. Chernyi. Present-day integrated information systems in chemistry and chemical engineering. (Creation of an integrated information system on VINITI premises. Working paper No. 1). Moscow, VINITI, 1970, 53 pp. (In Russian).
16. A. I. Chernyi. Integrated information system of Excerpta Medica Foundation. (Creation of an integrated information system on VINITI premises. Working paper No. 2). Moscow, VINITI, 1970, 43 pp. (In Russian).
17. A. I. Chernyi. Present-day integrated information systems in chemistry and chemical engineering. «Zhurnal Vsesoyuznogo khimicheskogo obshchestva im. D. I. Mendeleeva», 1971, 16, No. 1, 6—19. (In Russian).
18. A. I. Chernyi. Integrated information systems. Moscow, VINITI, 1971. 30pp. (In Russian).
19. Yu. Sheinin, E. Parnov, O. Simonenko. What are your working problems? (Questionnaire of the «Literaturnaya gazeta»). «Literaturnaya gazeta», Aug. 4, 1971, No. 32, p. 10. (In Russian).
20. V. A. Engelhardt. Integratism — the way from the simple to the complex in the cognition of life phenomena. «Izvestia AN SSSR. Seriya biologicheskaya», 1970, No. 6, 799—821. (In Russian).
21. V. A. Engelhardt. Integratism — a way from the simple to the complex. «Nauka i zhizn», 1971, No. 5, 8—15 (In Russian).
22. V. A. Engelhardt. The part and the whole in biological systems. «Priroda», 1971, No. 1, 24—33 (In Russian).
23. P. E. Rosove. Developing computer-based information sustems. N. Y., John Wiley and Sons, Inc., 1967, xiii, 384 pp.

P. J. Vinken
(The Netherlands)

THE ACTIVITIES OF EXCERPTA MEDICA

Excerpta Medica was founded in 1946 to face the challenge of the world's burgeoning biomedical literature. A group of Dutch physicians founded it as the nucleus of a reliable information service which would screen, select, index, abstract and disseminate, in the English language; the international biomedical literature in a most comprehensive way.

In those early days «a most comprehensive way» entailed the occupation of two floors above a shoe shop in Amsterdam's Kalverstraat. Now Excerpta Medica and its subsidiary groups occupy more than 10 buildings in and around Amsterdam, including a computer complex in the suburb of Osdorp.

To keep pace with the ever-increasing speed of information transmission, and to take full advantage of the developments in the nascent computer industry, Excerpta Medica established Infonet. Infonet is a fast-growing division which spearheads all the progress in automation and computerization. It designed and developed the automated system for the storage and retrieval of biomedical information which was first operated by Excerpta Medica in 1968.

Before explaining Excerpta Medica's system in detail, it can be indicated how considerably that system has evolved during the past 24 years by examining the existing Excerpta Medica/Infonet structure, and the shape of things to come.

Infonet has daughter companies and divisions specialising in different areas of study:

— **Dynaflow**, which handles the applied mechanics, including software for experimental and theoretical research and calculation in heat transfer, fluid dynamics, and one- and two-phase flow systems.

— **Digityp,** offers automated typesetting, page make-up and software for book production. It is the graphical systems division.

— **Rescona Engineering**, is an engineering consultative company which invents new automated production processes, instruments for hospitals, etc.

— **Infonet Electronics** is the hardware division, adapting computer equipment and peripherals to meet specific requirements; it is active in data engineering, cybernetics, etc. New equipment is developed where adaptions do not suffice.

— **Eltrac** is the division which anages the Excerpta Medica computer center at Osdorp, in Amsterdam. It is staffed by systems experts, programmers, and hardware engineers.

The center is based on an NCR 315 RMC processor, with its internal memory extendable to 160 k and a very large magnetic card storage in the form of CRAM-5 units (some units are already reserved for Infonet's planned on—line networks, job scheduling, for batch processing with interchangeable decks, etc.; additional units, are available for other real time applications). To extend the memory even further, two magnetic tape units, each of 32 kc, were connected.

An NCR-321 switching computer has been installed to control on-line use of the databanks: up to 100 terminal connections are possible.

There are programs interlinking the hardware components of the configuration, input and output routines, retrieval operations — either for specific search requests or for the dissemination of current and/or retrospective information — and for the on-line use of the databanks. Infonet makes available a series of programs for subscribers to the Excerpta Medica computer tape services enabling them to make the most effective use of the tapes in their own computer facilities.

Infonet's new automated information networks

Integrated Hospital System; Infonet plans to take the strain off the hospitals in the Netherlands by automating the transmission of information within each hospital and by terminal-linking the hospitals in a national network. All administrative work, classification, patient records, files of urgent data, and medical research will be streamlined.

So far, Infonet has linked its computer centre with a visual display terminal at Leiden University Hospital. It has proposed the integration of medical registration files. Work is now in progress to interlink the archives of the pathological institutes of Amsterdam, Leiden and Rotterdam, and the TNO Institute at Rijswijk. (see figure 1 and 2).

Integrated Libraries System; (see figure 3) a group headed by the director of the Royal Library at the Hague is assessing a system planned by Infonet for an automated Dutch national bibliography and documentation depot.

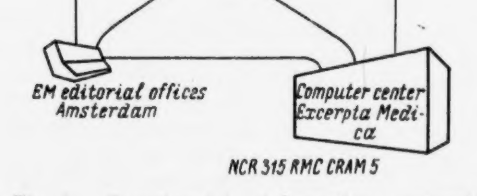

Experimental Information Network ›EM/CRI project‹
IBM 360/50 Data cell
Computer center University Leyden
Academic Hospital Leyden
EM editorial offices Amsterdam
Computer center Excerpta Medica
NCR 315 RMC CRAM 5

Fig. 1. Experimental information network «EM/CRI project»

This system — the Pilot Project for Integrated Catalogue Automation (PICA) — proposes a new central catalogue which will be consulted by scientific, university and general libraries in the Netherlands and abroad by telecommunication or interchangeable computer tape.

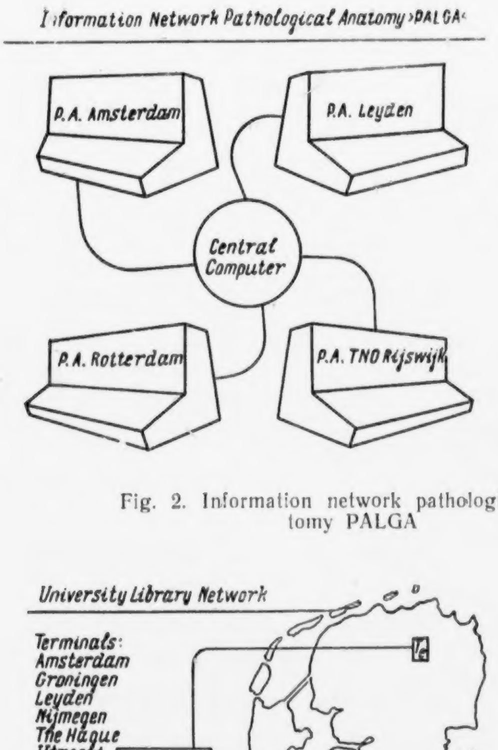


Fig. 2. Information network pathological anatomy PALGA

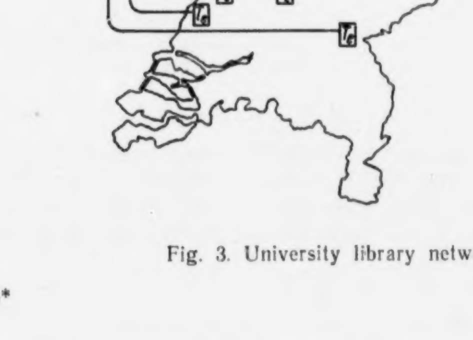


Fig. 3. University library network

PICA will be compatible with other library systems such as the MARK II but unlike other systems under development, which employ sorting techniques, Infonet has devised a search pattern allowing for inter-computer conversation. This will enable the library user to formulate the most precise definition of the publication which he seeks, by author, title, subject, form of publication etc. Infonet's policy in constructing an information system calls for the most intensive co-operation of software and hardware experts on the one hand, and the key specialists in the pertinent field of information on the other.

This ensures that those who will eventually be working in co-ordination with the system have participated fully at all stages in the planning of that system.

The effectiveness of this policy was clearly illustrated in the development of the Excerpta Medica system, which was shaped by continual dialogue between Infonet designers and the staff of Excerpta Medica who now work with the system (see figure 4).

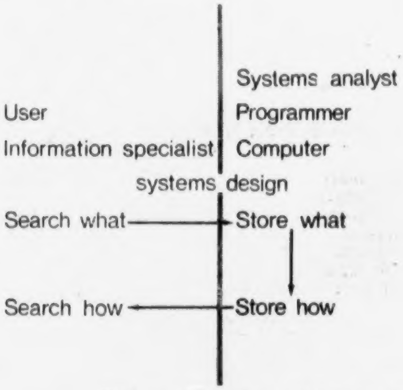


Fig. 4. Continual dialogue between Infonet designers and the staff of Excerpta Medica

Why Exerpta Medica?

The so-called information explosion has been with us a long time. We have had a rapid rise in the number of technical and scientific journals since the turn of the century. It is particularly interesting that those countries, such as Japan and the USSR, in which technological development had a later onset than the USA and Western Europe, are now experiencing a rapid growth in the production of technical and scientific information.

The biomedical area, with which Excerpta Medica is concerned, represents a very large proportion of the whole scientific literature field, a relatively constant proportion in all countries — contrary to the idea many people may have that most journals are technical or in the physical sciences. The linguistic problems with which Excerpta Medica is concerned — abstracting and translating the foreign biomedical literature into English — are increasing in importance rather than decreasing. Contrary to general belief, English is not becoming more and more widely used as the language of publication, but actually reached a peak, in terms of proportion of journals, in 1920, since when its position has been increasingly eroded by the widening use of Russian, Spanish, Japanese and languages of the «third world». Briefly, that is the scope of the problem which Excerpta Medica is tackling.

Excerpta Medica is probably best known in medical communities throughout the world for its publication of monthly abstract journals (see figure 5). In addition to the 36 journals of the regular series, into which the field of biomedicine has been divided, another 20 or so «special services» are published on more limited subject areas such as Fertility Control, Atherosclerosis, Contact Lenses, Gram-Negative Infections, Epilepsy, Multiple Sclerosis, and the like. All of the regular monthly journals and some of the special services are now «composed» entirely by computer, and all the information is stored in the computerized databank. Whereas the 36 journals of the regular series are distributed on subscription, the more highly specialized special services are sponsored by various pharmaceutical companies, societies or governmental organizations, and are therefore generally distributed free to a select mailing list. The computerized databank, provides bibliographies and answers to questions on demand. Excerpta Medica also publishes textbooks, symposia, monographs and the proceeding of congresses, and has assisted in the organization of symposia, provided translation services, and done developmental work in the audiovisual area (films and closed-circuit television).

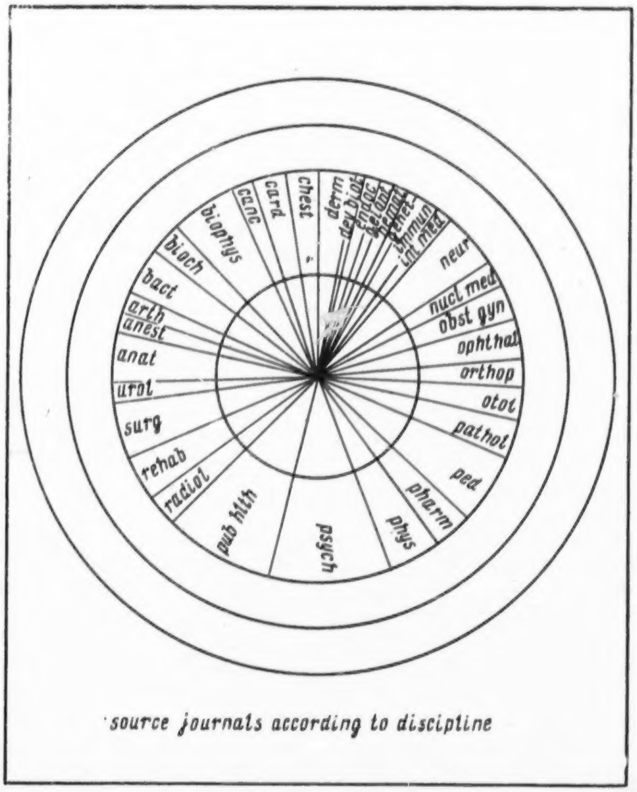


Fig. 5. Source journals according to discipline

About 500 full-time staff members work in the Amsterdam offices, plus the part-time physician-editors, in addition to the 700 editorial board members and some 4000 part-time abstractors who are scattered all over the world. More than 3500 biomedical periodicals, are received by Excerpta Medica from possibly 15,000 such periodicals published in the world; his results in a flow of about 100 issues per working day, which yields about 250,000 biomedical articles per year. All of these are classified, indexed, and stored as citations in the databank, and about 60% of them are abstracted (the most significant ones). Each abstract appears, on the average, in about 1.6 different Excerpta Medica publications. All of the journals received are microfilmed in collaboration with the Royal Netherlands Academy of Sciences, resulting in a microfiche file which now contains about 20,000,000 pages.

Each of the 36 regular sections has an international editorial advisory board, totalling some 700 physicians and other specialists in all, to collaborate with the executive chief editors and the editors of the individual sections to provide direction and guarantee high standards of quality in selection, abstracting and indexing.

Most sections have more than one editor and some have 3 or 4, resulting in a total of about 80 specialists. These specialists do all the selection and indexing and supervise the content of the abstracts. Excerpta Medica has a linguistic problem however which is aggravated by the fact that it publishes in English in a non-English-speaking country, so that most of the editors lack a native knowledge of that language. Consequently the linguistic supervisors (usually physicians of English or American extraction, living either in Amsterdam or in England) edit all texts before publication from a linguistic point of view. We also require a corps of translators to translate abstracts that have been prepared in other languages (often the original language of publication, sometimes Dutch or German) into English. The translation bureau, which was originally established to satisfy the internal needs, has now expanded into one of the largest biomedical translation bureaus in the world, providing about half of its 2,000,000 words per year to outside clients. One of the characteristic features of this service is that all work, about 70% of which is into English, is supervised by a second person with a native knowledge of the target language.

In order to separate the very best of the source material, the 3500 journals are sorted according to quality. Essentially, the world's biomedical literature is divided into 4 roughly equal groups of journals: general medical journals which may contain articles on any subject and which can be divided into «good» and «mediocre» general journals, and speciality journals (also divided into «good» or «mediocre»). The speciality journals are divided into the 36 current sections of Excerpta Medica. This information («General» or «Speciality», «Good» or «Mediocre») is coded in the databank, **a priori** per journal, so that a searcher can restrict his search to a limited segment of the biomedical literature if he so desires. Excerpta Medica also provides partial coverage of a number of general scientific journals, such as Nature and Science, which frequently publish articles of a biomedical nature.

Since there is some inconsistency in the use of the term «citation» as it pertains to a scientific article, the elements of the citation as used by Excerpta Medica should be defined. For every article, there is an English title and (unless written in a different orthography) the foreign title, codes for the language and country of publication, the first three or four authors, the author's institutional address, the abbreviated name and Coden of the periodical, the volume and issue number, pagination and year, and the year of input into the databank (this is given by the first two digits of the production number).

Processing of source journals

All of the source journals are first microfilmed, cover to cover, in conjunction with the Royal Netherlands Academy of Sciences, and are stored as microfiche. The journals are then brought to the editorial department where they are first examined by the assignment editors. These assignment editors make a preliminary assignment of each original article, including editorials and letters to the editor, to one or more of the 36 sections of Excerpta Medica; thus, by writing the section number or numbers on the article, they provide what we can call the first decimal of the classification. The journals are then made accessible to all of the section editors for 5 days, during which time they can add their section numbers to other articles which may not have been assigned to them originally. This 5-day period also gives the editors of the special services (sponsored SDI journals) the chance to select articles for their publications, which are intended to be reports of only the «cream» in their selected fields. From the preliminary assignment stage, the journals are sent to the coding department where the citation of each article is prepared for input. This means that appropriate code letters are assigned to the author's name and address, original and English title, etc., that the appropriate abbreviations are indicated, and that a correctly edited English title is provided where necessary. This information is then punched onto paper tape by Flexowriter and fed into the computer with the section numbers. The computer provides an immediate print out of each day's input which is then sent to the proofreading department along with the original journals. Any corrections necessary are made directly in the computer memory by means of a correction program, after which the citations are permanently entered into the databank and assigned a production number. On a weekly basis, the computer prints a series of composite forms, the abstract-index form, for each citation and each section to which it has been assigned; each part of this form contains the full citation of the article plus the section number. The journals are then torn apart and each set of abstract-index forms is clipped to the corresponding article.

These articles with abstract-index form attached are then sent, in turn, to the classification and index editors of each interested section. The chief editor of each section, beginning with the primary section, must do two things: classify the article into one or more of the subcategories of his section, and decide whether he wants an abstract of the article to appear in his monthly booklet. The article is then passed to the index editor (usually a different person), who assigns the primary, secondary and item-index terms. As soon as the article has been classified and assigned primary indexing terms by the first section editor, the appropriate form is torn off and sent to the computer, before any work is done on the abstract. The composite form provides a separate detachable sheet for classification numbers and primary indexing terms for each section so that these may be written and entered into the databank without awaiting further secondary indexing. If a search is done while an article is still being processed, the searcher at least has a chance of retrieving a classified and partly indexed citation, and later a fully-indexed citation, while the abstract is still in preparation. The abstract form is then sent, with the article, to the editor of the primary section, who is responsible for preparing the abstract. If the primary section editor feels that the article is not worth abstracting, then the editor of the next section receives the responsibility. Assuming that at least one editor wants an abstract for his section, he sees to its preparations either by writing an abstract himself or by having the article sent to one of our 4000 medical abstractors all over the world. In many cases the abstract will then have to be translated into English, and in most cases the English version will have to be checked by a linguistic supervisor. Eventually, the entire abstract is also fed into the databank.

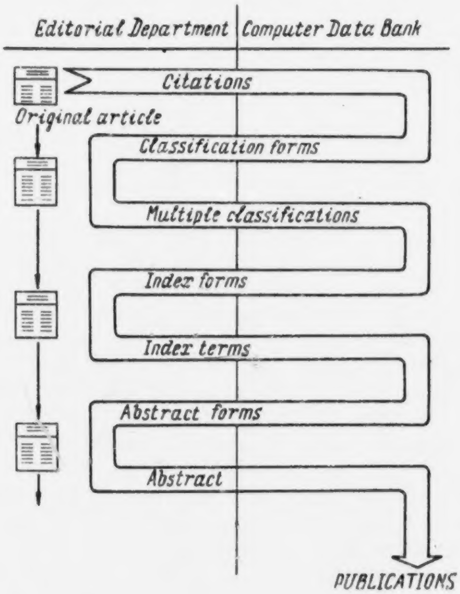


Fig. 6. Processing of articles for input of appropriate information into computer data bank

Although the classification and indexing is done individually for each section, the same abstract which has been prepared for the primary section is used, with a few exceptions, for the other sections. At any stage, citations with or without abstracts can be printed out, as hard copy or on magnetic tape, in response to search requests. When sufficient abstracts have been accumulated in a section for the monthly journal, the computer automatically composes the pages, assigns abstract numbers and page numbers, compiles the author and subject indexes, and produces a film which can be sent to the printer. One of the characteristic features of the Excerpta Medical operation is the constant interaction between the computer and the editorial staff. This is an example of a completely integrated information storage and retrieval system.

An information system is only as good as its index, which represents the handles by which the information can be retrieved. If the indexer and the searcher were always the same person, then almost any indexing system would suffice, but the problem is precisely that the minds of the indexer and the searcher, even when both are medical specialists, often run in different channels. For this reason, the editors of Excerpta Medica have developed an indexing system which works by means of three different approaches: the subject index, the classification system, and the item index. The classification system and item index are **a priori**, fixed or semi-fixed systems, while the subject index, which provides the most specific retrieval, is a free **a posteriori** system which has two levels — the primary terms and the secondary terms. All indexing in Excerpta Medica is done by medical specialists, and the goal is to provide access on either a broad or highly specific basis, according to the needs of the searcher (see figure 7).

DOCUMENT PROCESSING →

Subject Indexing	Classification	Item Indexing
Medical specialists	Medical specialists	Medical specialists
Open: free choice of words	Open ended	Closed
Two levels of weight	Decimal	999 possibilities
Automatic permutation of primary terms	Polyhierarchical	
Automatic thesaurus		
70000 primary concepts	2000 categories	50 categories
Depth: no limitations	Maximum depth: 5	Maximum depth: 10

Fig. 7. Document processing

The editors of each section have specified a detailed classification system with a possible depth of 4 decimals, so that the various subdivisions of all the sections total about 3,000 pigeon-holes into which any article can be placed. This system is open-ended, so that new subclassifications can be created at any time, and is polyhierarchic in structure. A maximum of 5 different classification numbers can be given to any article within each section; of these, the first determines where the abstract will appear in the monthly booklet, but all are equally valid for retrieval purposes. Retrieval by means of classification number is a very powerful tool in many cases in which the search question coincides with a subcategory.

The item index is a list of terms which are considered to be too broad for use as primary index terms but which are often useful for specific retrieval. These are comparable to secondary subject-index terms, which are not thesaurus-controlled and are therefore generally not computer-searchable, to which numbers have been assigned. The item index is a controlled list of secondary concepts, with a maximum of 999 numbers, only about 55 of which are now in use; up to 10 numbers can be assigned to an article. These are mostly concepts such as «child», «in-vitro study», «rat», «model» which serve to specify a search further in conjunction with the appropriate subject-index terms or classification numbers. The item index is intended purely for computerized retrieval and does not appear in the monthly journals.

The goal of Excerpta Medica is to index as specifically as possible, much more specifically than woud be possible with a limited number of terms, and to make maximal use of the computer while keeping the intellectual work in the hands of the medical specialists. The subject index consists currently of about 120,000 thesaurus-controlled primary terms — each of which is stored in the databank in the form of a 7-digit number, plus see — also references from broad terms to narrow terms — and about 250,000 synonyms. These synonyms are also stored in the databank (the thesaurus file or MALIMET), so that the indexer or searcher need not consult the thesaurus but can use natural language to express his ideas, language which the computer will translate into accepted preferred terms. There is no limit to the number of index terms which can be assigned to an article. The indexer freely assigns any secondary terms which may occur to him, sometimes forming a sort of telegraphic mini-abstract.

When it comes to printing the subject index in the monthly journals, the primary terms are automatically rotated so that each appears in its proper place in the alphabet, followed by the other primary terms and then by all the secondary terms. Thus a search «by hand» for any of the preferred terms will unearth all the other related indexing terms. Significantly, all indexing at Excerpta Medica is done from the entire source article, not just the title or the title with the abstract.

The synonym problem, a serious stumbling block for any free indexing system, has been solved at Excerpta Medica by the development of MALIMET, the Master List of Medical Indexing Terms, a computerized thesaurus which is far too comprehensive to be used by hand (the computer print-out is a pile about two meters high) but which automatically translates synonyms into preferred terms. For example, the word «cerebral tumour», in the British spelling, is automatically transformed into «brain tumor» by the computer. Only «brain tumor» will appear in the journals, and only the numerical equivalent of «brain tumor» will be stored in the databank. MALIMET was originally built up by asking each of the section editors to provide a list of primary indexing terms for his field, along with as many synonyms as possible. After eliminating the internal contradictions, these 25,000 or so primary terms were then fed into the computer with their synonyms. During the next two years, whenever the computer was offered a term which it did not recognize, this term appeared on an «error list», which was then processed by the specialist index editors. Every term on each weekly error list had to be disposed of either by entering i as a valid new primary term, or as a synonym of an existing preferred term (so that in the future it would be translated automatically), or by correcting a typographical error. In this way (and the process is still continuing), the number of preferred terms grew to about 120,000 and the number of synonyms to about 250,000. The only type of preferred term still being added in significant numbers are drug names, but the synonym total is expected to go about 500,000.

Examining the organization of our computer files (see figure 8) in a little more detail, let us first look at the two files required for our thesaurus: an output file in which the preferred terms are listed in numerical order, and an input file in which all preferred terms and synonyms are listed in alphabetical order, followed by the preferred term (PT) number. How does this operate in practice? Any term chosen by the indexer is transformed into the proper number and stored in the databank in numerical form. During the retrieval process, the search terms are again translated into numbers, and the numerical then transformed into analphabetic output by means of the output file.

Thesaurus file organization

output file 1: PT number to PT

0000001 ····▸ PT_P
0000002 ····▸ PT_K
0000003 ····▸ PT_A
0000004 ····▸ PT_E
0000005 ····▸ PT_M
0000006 ····▸ PT_H

input file 2: Any term to PT number

PT_A ▸ No A ▸ A_1 A_2 A_3 see also..
PT_B ▸ No B ▸ B_1 B_2 B_3 see also..
Syn. A_1 ▸ No A ▸ PT_A
PT_C ▸ No C ▸ C_1 C_2 see also...
Syn. B_3 ▸ No B ▸ PT_B
Syn. C_1 ▸ No C ▸ PT_C
PT_D ▸ No D ▸ D_1 D_2 D_3 D_4 D_5 see also.

Fig. 8. Thesaurus file organisation

All of the information in the databank is stored in three different levels, each representing a different degree of accessibility. The first-level or rapid-retrieval file is characterized by its short, fixed-length terms, permitting sequential organization and rapid access. This file contains the primary and item index terms (which, of course, are stored in numerical form) plus the classification numbers, Coden code and year of entry, and finally the «address», or coded location, of the rest of the citation and abstract data in the other files. A search of this file is rapid but can only yield numerical information, which must then be used as the key to obtaining the entire citation or abstract from the second and third levels. A search for a particular author, since this is not a fixed-length term, is thus slower than a search by primary index entry (see figures 9 and 10).

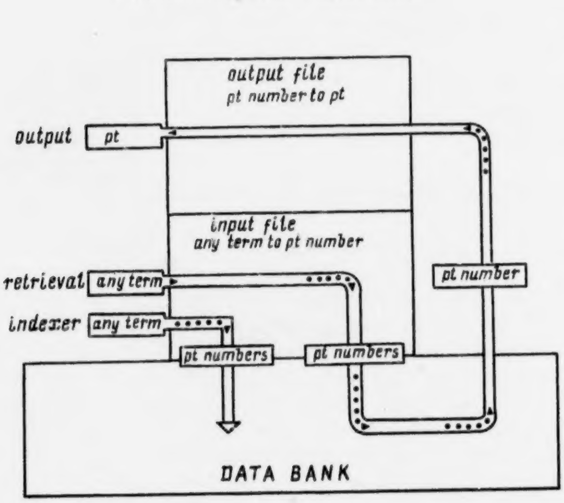


Fig. 9. Thesaurus: system's supervisor

The organization of the files can be explained in a slightly different way as follows. One magnetic card is devoted to a sequential address file listing production numbers together with the «address» of the citation, index terms and abstract for that production number in the other files. Another file is a sequential listing of citations in numerical order (since the citation is fed in at the time that the number is assigned); these citations, however, are variable in length. Another file consists of production

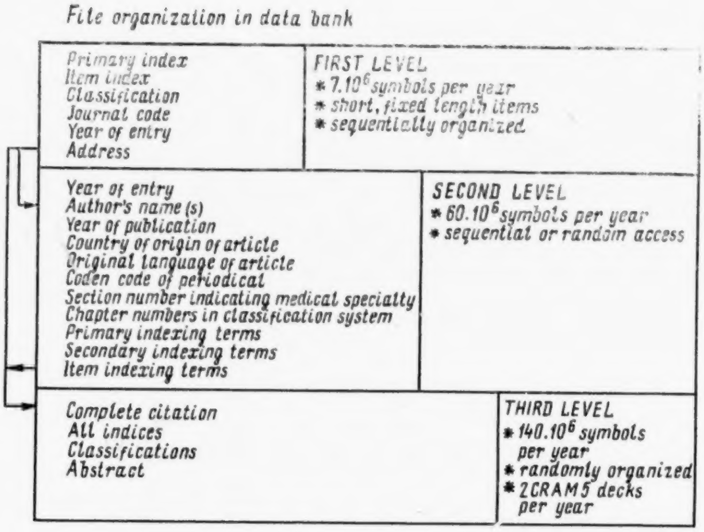


Fig. 10. File organisation in data bank

numbers, listed at random, each followed by all the primary index terms (in numerical form) and then by all the secondary index terms, and a final file consists of random production numbers each followed by an abstract. Note that the index file is not inverted, providing for much greater economy of operation when a large number of searches are run simultaneously.

The drug literature service or Drugdoc (see figure 11) is run in many ways just like any other section of Excerpta Medica. The section is very large, however (representing as much as 25% of the whole databank), and no abstracts are written specifically for it (although many of the articles may be abstracted for other sections). It is intended primarily as a computer service (although a printed, abridged publication called the Drug Literature Index is available). Input into the system for the drug project is the following information pertaining to every drug and chemical compound cited: the full citation of the article; the generic name of the drug or compound;

the experimental name and/or code, if an experimental drug; the trade, or brand, name; the chemical name of the drug; chemical structure information (Wiswesser Line — Formula Chemical Notation); clinical indications and contraindications; clinical and pharmacological effects; untoward drug effects; the name of the manufacturer and the location of the manufacturer. This information is extracted from 40,000 articles a year and more.

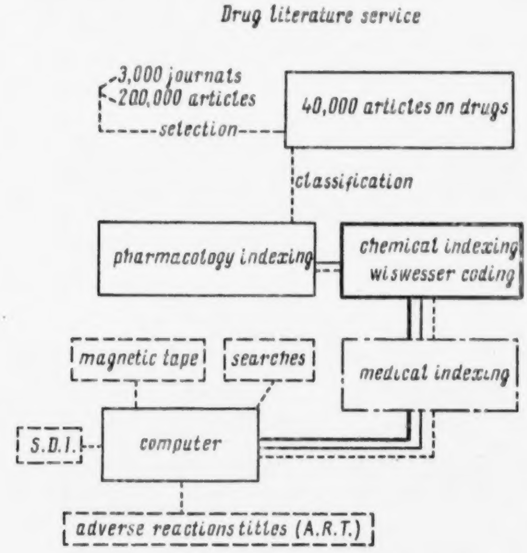


Fig. 11. Drug literature service

The articles are indexed particularly deeply from three points of view by three sets of specialists: medical, pharmaceutical and chemical. Many chemical journals have been added to the collection of biomedical and related journals to afford better coverage in this section, the purpose of which is to include articles describing biological effects of chemical agents, particularly drugs or potential drugs. These articles are classified into 186 pharmacomedical subgroups, and are then indexed by a physician, a pharmacologist, and a chemist who also provides a Wiswesser code. For an even more comprehensive documentation of drug literature it has been decided to use a chemical fragmentation code too. The information is stored together with all other Excerpta Medica information in the databank, where it is available for retrieval in response to search requests, and is also available in the form of regular computer tape and/or hard copy.

A special sub-section of the drug project is the Adverse Reactions Titles, a monthly listing of indexed classified titles for all articles reporting a side-effect of drug therapy.

A review of the databank status in June, 1970, showed that 213,369 citations for the 1969 series had been processed thus far; 91.7% of these had been classified and 77.5% indexed. For the 165,277 classified and indexed items, the average number of classifications per citation was 2.18 and the average number of primary indexing

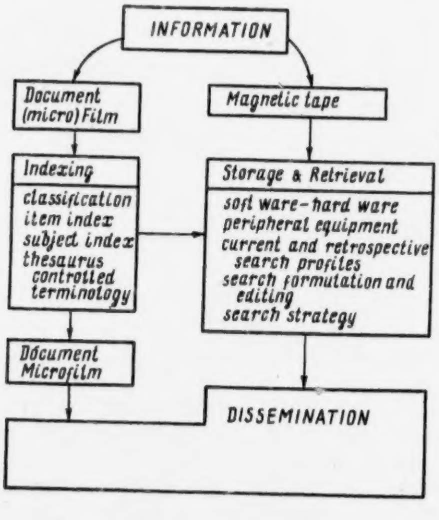


Fig. 12. Excerpta Medica's information dissemination scheme

terms per citation was 4.96; 135,183 of these items were abstracted. With particular regard to the 32,776 drug items in the series, the average number of specifically Drugdoc classifications and Drugdoc primary indexing terms per citation is 1.94 and 6.05, respectively. Half of all items were classified and indexed within about 12 weeks after input of the citation, while 50% of the abstracting effort was completed in about 22 weeks. (see figure 12, Excerpta Medica's information dissemination scheme).

W. Piróg

(Poland)

THE USE OF COMPUTERS IN SCIENTIFIC, TECHNICAL AND ECONOMIC INFORMATION IN POLAND

The problem of using computers in information processing have, for many years, been of vital interest to the Polish information service. It is the concern of The Central Institute for Scientific, Economic and Technical Information, a number of branch and plant information centres as well as of some universities and research establishements. Both theoretical as well as experimental and implementation works are now in progress. Special attention has been given to this problem by the Branch Centre of Technical and Economic Information at the Institute of Ferrous Metallurgy in Gliwice, Branch Information Centre at the Central Construction and Research Bureau of the Shipbuilding Industry in Gdańsk, Chair of Economics, Organisation and Planning at the Technical University in Wrocław, Information Centre of Electronics and Teletechnics UNITECH in Warsaw, Institute of Mathematical Machines in Warsaw, Institute of Building Technology in Warsaw, Ministerial Centre of Technical and Economic Information in the Ministry of Heavy Industry and National Library.

The forerunners of the presently developing initiative in the use of computers for information processing were the works accomplished as early as in the first half of this decade by the Branch Centre of Technical and Economic Information at the Institute of Ferrous Metallurgy in Gliwice, Branch Information Centre at the Institute of Zootechnics in Kraków as well as Chair of Economics, Organisation and Planning in Technical University in Wrocław.

The first two centres worked out and implemented the systems with perforated cards computers used, but the Chair used for its works the computer «ODRA 1003» constructed in Poland.

The research work and the tests that have, for several years, been carried on by the Information Centre at the Institute of Ferrous Metallurgy in Gliwice, have resulted in a peculiar system of information retrieval by means of perforated card machines of the Hollerith type installed in the specialized Entreprise of Analytical and Perforated Card Machines for Metallurgy.

Machine cards have been used to organize the holdings of all useful bibliographical units inserted in «Express Information of Ferrous Metallurgy and Steel» published by the above-mentioned Centre. A glossary of most frequently occuring and useful terms (2 500 words) and a code interpreter have been prepared for the description of documents. The cards are chosen from the collection of headings by means of a selector. For these headings tabulagrams are printed with the help of a tabulator that contains the data enabling to indentify the document.

To mechanize partly the information processes in the Institute of Zootechnics a set of perforated card machines of the ARITMA type and teleprinter have been used. The characteristic feature of the system used was also the preparation of a glossary of key words containing, in an encoding form, a set of chosen concepts used to describe the document. As a matter of fact, the glossary appears to be a classification system. The efforts that have been made to use the UDC symbols have not proved fruitful because of the insufficient depth of the UDC classification in the class 639 from the viewpoint of modern zootechnics.

As far as the patent information is concerned the Technical University in Wroclaw has carried out the experimental works for the Polish Bureau of Patents, Petrochemical Works at Plock as well as for the Union of Man-Made Fibres. The machine ODRA 1003 which has been used for these experiments does not possess an external memory and therefore could neither be used to form large information collections nor for automatic processing of them, but it could be used as a selector of recordings to be put down on the paper tape for making up the simple lists. The necessity to apply as a memory the perforated tape and organization of the tape files prevent from using this method for practical purposes. Nevertheless, the experimental works are being continued on information searching from the collection of patent specifications provided with the American patent classification.

Bibliographical descriptions of patent specifications are being carried on to a perforated tape. The program that is introduced into the machine contains, among other things, the table of conversion from the American classification to the classification used in Poland. . The machine reads the perforated tape, and descriptions which correspond to a given class used in Poland are written down on the printing device in the form of lists.

The next centre of the initiative taken on the use of computers for information processing is the shipbuilding industry — The Central Construction and Research Bureau of the Shipbuilding Industry in Gdańsk which cooperates with the Branch Information Centre. For these purposes the computer ELLIOT 803, equipped with the tape memory, was used, and the system was given the name APIS (Automatyczne Poszukiwanie Informacji Syntetycznej — Automatic Search for Syntetic Information). The system was elaborated in 1967 and put into operation in 1968, and is designed:

1. to make up and keep up-to-date the collection of bibliographical information
2. to prepare the topical bibliographical lists on the requests of users.

The information is a bibliographical description marked with the appriopriate heading of their own branch decimal classification. Bibliographical descriptions are put down on the card Format A6, and are carried on to a paper tape by means of the automatic writing device «OPTIMA 528». This tape constitutes the collection which brings up to date the contents of the basic collection and, besides' it contains the subject matter of the next issue of the Information Express — BIS published by the Centre. The primary task of the APIS system is to make up topical bibliographical lists, on the basis of the existing holdings to meet the user's requests. The user's enquiry is being transformed by a specialist from the section concerned with cooperation with a computation centre, in accordance with the rules and terminology for the searching program. At the same time twenty inquiries can be introduced into the machine. The program enables to restrict the topical bibliographical list to a definite period and definite journals and enables to observe these two limitations. It also makes possible to print the topical bibliographical lists according to the heading words of branch classification, and after the names of the ships.

The information processing in the APIS system is performed by means of the five programmes: updating (making the main collection more complete), the control printout of a complementary collection, making up of the topical bibliographical lists on the ground of searches done in the main collection, modification of the main collection and subsidiary collections, bringing up-to-date the main collection by removal of the outdated items from it and from the subsidiary collections and the printout of these items.

The subject of the continued research done by the Central Construction and Research Bureau of the Shipbuilding Industry in Gdańsk is to organize, by exploiting computer for information retrieval relevant to the factographic data relating to technical parameters of ships, data needed for designing calculations, building technology as well as the information for menagerial staff. The research and experimental works are now in progress.

Much attention is given to the automation of information processes by the Institute of Mathematical Machines in which the Branch Information Centre develops its activities. In 1967 the Institute undertook research and preparatory work with the aim to mechanize the publishing works of the «Information Bulletin» of the Institute of Mathematical Machines — Bibiliography, elaboration of the subject index, and information retrieval. The subject of the publishing work on the «Bulletin» are bibliographical descriptions relating to computers, and for marking them with their own classification has been adopted. The system has been called INBI (Informacja bibliograficzna — Bibliographical Information). The first version of the system has been accomplished by means of the Polish make machine ZAM-21 which took the name INBI-A. Because of untypical classification used and its unique nature, the system could not be widely made known.

The next version of the INBI system is the INBI-B worked out for the Polish machine ZAM-41 which is equipped with the units of magnetic tapes. The bibliographical descriptions for this system were enriched with the additional details relating to the date of publication, language symbol and the symbol of the kind of a document. The INBI-B system is of the publishing nature. The system of searching programmes has been elaborated for the retrieval of bibliographical information which was given the name INBIS (Search for Bibliographical Information). The system also functions by the use of the machine ZAM-41. The selection of the information from the collection and its preparation in the form of topical bibliographical lists can be made to answer enquiries which correspond to the classification part (symbols' part) of the description or to their textual parts. As regards the last case, descriptions are chosen according to the correlation of word sequences contained in an inquiry with those in the collection. A search request formulation is introduced into the machine with the help of a perforated paper tape. The time of search done in collections depends on the degree of complexity of requests. The time of searching of 1000 items in the collection for simple requests with the classification part amounts only to about 1 minute, but for complicated requests with a widely build textual part the searching time of 1000 items in the collection is 7—8 times as the former. The IBIS system is, undoubtedly, very interesting, but it could not be widely used, and therefore it ought to be considered experimental as far as the recording and alphanumeric search are concerned.

The problem of mechanization and automation of information processes were the subject of interest to the Central Institute of Scientific, Technical and Economic Information as early as the years 1964—65—66. The job that was done was of introductory nature and it was limited to the research on literature and elaboration of the idea aimed at organizing and carrying on the experiments. Some works being the part of the research problem were, among other things, carried out within the compound research plan which was worked out by the

Standing Working Group for Scientific and Technical Information acting within the Council of Mutual Economic Cooperation (SEW). They may be regarded as theoretical, being concentrated on the classification and information retrieval languages, research on different kinds of indexes, etc. These works are continued. Special attention is being paid within this group of problems to descriptors and thesauri as well as to the UDC problems. In 1967 it was resolved by CIINTE to work out the experimental system for one of the central information holdings run by the Institute. For these purposes the central register of research papers already completed by the Institute was chosen. On the basis of this publication «Guide of Scientific and Research Papers» was coming out quarterly. Decision was taken to prepare this reference publication by means of a computer, as the first stage of automation. The cooperation was established with ZOWAR (Zakład Obliczeniowy — Computation Centre) in Warsaw which disposes of a computation centre and specialists-programmers. Then it came to set up a group of people that includes the specialists from CIINTE and ZOWAR, with the aim to work out the project of the system. The order for elaboration of appropriate programmes has been placed with ZOWAR. The most characteristic feature of the system is the use of UDC for the grouping of information in the above-mentioned publication. The system is called IGA (Informacja Grupowana Automatycznie — Information Grouped Automatically) and programmed to be realized by the computer IBM-1440 (machine with disc memory). About 50 programs were elaborated for the system providing it with the supervising program, thus enabling the automatic control between the programmes to operate. The publication contains a collection of data arranged according to the UDC symbols, the subject and the author index and also a list of contents. The IGA system came into operation in 1968, undergoing different mutations. At present «the Guide» comes out every month and in includes the bibliographica descriptions of the completed research papers and the above-mentioned indexes. The work has now been started on the reprogramming of the IGA system for a tape computer as well as on the programmes designed for searching the required information in a collection.

Another initiative taken by CIINTE was the elaboration of a project of the system to make up, with the help of the machine ZAM-41, the KWOC index intended for the collection of documentation cards relating to information problems run by CIINTE. Basing on this collection CIINTE has, for several years, been publishing «Przeglad Pismiennictwa Zagadnien Informacji» (Review of Literature on Information Problems). In order to project a system a group of specialists from CIINTE and from the Institute of Mathematical Machines was organized by the end of 1967. The project was completed, experiments were conducted, and the system was put into operation in the third quarter of 1969, and as early as 1967 a reference book consisting of two parts (main bibliographical list and KWOC index) was published. Each item of the bibliographical list is referred to the corresponding item of the PPZI issue. When accomplishing this project the main collection of key words was build up. The system can be utilized to make up the KWOC index for other collections of documentary information run by CIINTE.

In 1967 CIINTE undertook the initiative in the development of cooperation with branch information centres on the preparation of systems, making the experiments concerning the use of computers for publishing processes and information retrieval. Furthermore, CIINTE put forward the idea of organizing, at separate computer centres, the teams of the branch centres concerned called «the clubs of computer users». In 1968 cooperation was undertaken with the Information Centre of Electronics and Teletechnics UNITECH which took the initiative as early as 1967 to utilize the computer for the works connected with the publishing of «Current Bibliography» coming out once a week, and for the search of information in the collection. For the description of documents UNITECH prepared their own glossary of descriptors which is currently brought up to date. As a result of cooperation between the Information Centre UNITECH, CIINTE and the Bureau of Research and Projects for the System of Electronic Data Processing, the printing and retrieval system was accomplished for the machine ICT-900 called by its cryptonim ASIA (Automatyczna Selekcja Informacji Adresowane — Automatic Selection of the Addressed Information). The bibliographical description made on a special form is, by means of the automatically writing device «Optima» 528, carried on to the eight-channeled tape perforated according to the ABC code. This tape operates as a carrier of data to be introduced into a computer. At the output of the computer, for preparation of the column «Current Information», a perforated tape according to the «OPTIMA» code is being obtained. The recording of the text is performed by the automatic writing device «OPTIMA» 528. The ASIA system, in its printing part, has already been put into operation. The research on the elaboration and implementation of the programmes for retrieving the information requested have also been completed.

In addition to UNITECH, other introductory research work connected with the use of a computer in information processes is also, for several years, conducted in the Ministry of Engineering Industry and Ministry of Heavy Industry to meet their own needs. The main stress was laid on the elaboration of foundations suited to a new form and contents of bibliographical description by a wide application of descriptors. The problem of the application of a computer for information processing was the subject matter of training courses in the form of conference. A number of branch centres are working on this problem. These are: the above-mentioned Branch Centre at the Institute of Ferrous Metallurgy in Gliwice, Branch Centre at the Industrial Institute of Agricultural Machines in Poznan, etc. In 1967, to develop the research work the Centre of Technical and Economic Information Processing was established under the Ministry of Heavy Industry. In 1969 the Centre worked out for the machine ICT 900 a project of the publishing and retrieval system ASIW to handle the information relating to business trips made on the initiative of the Ministry of Engineering and Ministry of Heavy Industry. This system has already been put into operation. The above-said centre is also preparing the foundations for working out of the SAGO-CZAS system for the machine ICT 1900 in order to print and retrieve the information about articles in scientific and technical journals. In both systems, subject headings, in addition to the bibliographical description, were used to describe the content of documents.

In the years 1969—1970, CIINTE, in cooperation with the Institute of Building Technology worked out the project of the automatic retrieval of patent information (SAWI-2P) for the machine Minsk-22. For experimental purposes a collection of patents within the class 37 (building structures) and 80 (building materials) was prepared. The aim of the system is to solve the problem of retrieving from the information collection contained in the memory of a computer the necessary patent specifications with the help of its various data (different classifications, formal and legal elements and descriptors).

An introductory work on the elaboration and publication of catalogues and current bibliography has been done in National Library which cooperates with the Institute of Mathematical Machines and utilizes CIINTE's experience.

The initiative taken in the field of research, design and experimentation to use a computer for information processing has, for the last two years, extended and included new centres and groups of specialists. The present condition may be characterized as follows:

1. So far, the job taken and works on the use of computers for information processes in Poland ought to be considered as introductory to make the groups of specialists familiar with the problem and to train the staff
2. The experiments conducted will enable to collect experiences which will help to work out the projects fitted for a wide application of computers in the national system of information, bearing in mind the nationwide automated system of information
3. The present research and experimental works have been carried on in a dissipated way, in different conditions, with different equipment used, and devoid of a common aim and uniform foundations. Different carriers, of information, different classifications and indexing languages have been used. They may then be regarded as introductory and experimental as compared to the advanced stage of the automated system of scientific information in the world.

The forecasts relating to the development of scientific and technical information for the next 5 years provide the concrete and successive implementation of automated processes for the nationwide system of scientific and technical information. The fundamental requirements to be put on the system of automated information is to ensure the universality of information flow. It results from the designed and realized organizational structure of the country information service adapted to the system of the planned economy.

In the present manual system of information the university flows is assured. In the automated system it makes necessary to standardize the documentary entry, indexing language and parameters of the equipment and outfit. The starting point is to coordinate research and experimental works. These works, in compliance with the plans worked out by CIINTE and the Plenipotentiary of the Government for Electronic Data Processing, will be developed in the field of automated printing and searching systems for various sciences, in cooperation with the information centres concerned. As the present forecasts indicate, in the years 1971—1973 a project of national-wide system of automated information will be worked out, and in the years 1974—1975 and in the years to follow gradual implementation for practical purposes will be performed. The works that deserve the greatest attention are those relating to the elaboration of a subsystem project for the automated patent information being run by the Bureau of Patents and CIINTE, the works on the automation of printing processes of central catalogues, catalogue publications, current bibliography to be done by The National Library, as well as the works on the automation of information in the field of specialized holdings done jointly with the Centre of Economic and Technical Information Processing of the Ministry of Heavy Industry and Ministry of Engineering Industry. At the same time, alongside of the experimental works it is scheduled to conduct the experimental works.

The integral part of the system designed will be also the subsystems of specialized central holdings now organized by CIINTE. They consist of information holdings concerning the completed research papers, translations of foreign scientific and technical publications made into Polish, topical surveys already prepared, publications on information problems, documentation and library science. The important enterprising will be the subsystem of current awareness information relating to the articles in scientific and technical journals. The important element of the system is the central computer centre to be set up at the end of the next 5 years in order to make possible servicing the central holdings of information, centralization of information publications of the current awareness type.

The elaboration and implementation of the project for automated information on the nation wide scale is not limited to the initiative taken in the preparation and realization of fractional projects relating to individual fields. However, it is assumed as a condition to harmonize the efforts with the coordinating institution — CIINTE. The essential condition for the elaboration of the nationl-wide project is the development on a large scale of research and experimental works. There are difficult problems among them to be solved: classification systems, indexing and information retrieval language. Efforts are to be made to work out a universal information retrieval language.

The project in progress will be harmonized with the project of international system for scientific and technical information in the countries grouped in the Council of Mutual Cooperation which is being worked out by the International Centre of Information in Moscow. In this case the coordination is indispensable to ensure appriopriate cooperation.

The project should also take into account the world trends in mechanization and automation of information processes. It is necessary to present some of them as the most important.

- setting up of large branch automated centres of information (printing and retrieval systems) capable of providing the users with the encoded information tapes
- automatic processing of documents (abstracting)
- input of many texts into the memory of a computer by rapidly working readers of texts
- storage of a great number of informaticns in the external memories of computers
- utilization of a computer through the transmission data network

- elaboration and application of the devices for storage and printing of graphical information,
- receipt of information from a computer immediately on to microfiches and mirofilms
- searching in microform collections by computer
- working out of a modern retrieval language with universal nature

In the light of the debate held at the Second International Semimar on the UDC and Mechanized Information Systems (Frankfurt on Main 1—5 June, 1970) the UDC could be such a language after appriate modification.

The introduction of computers to the information system requires of a qualified staff to be able to carry on research and to utilize the automated subsystems which are put into operation.

REFERENCES

1. W. Klepacz and D. Prawdzic. System informacji bibliograficznej INBI (Komunikat niepubl., zgłoszony na sympozjum w Moskwie w 1967 r. The INBI System of Bibliographical Information (the unpublished information presented at the symposium in Moscow, 1967)
2. W. Piróg, F. Kościński and M. Mikołajski, Wykorzystanie maszyn cyfrowych w służbie informacji naukowo — technicznej i ekonomicznej. The use of the computers in the service of scientific, technical and economic information. Orga. Samorz. Zarządz., 1968, № 3 pp. 113—120
3. H. Richter, Krótka charakterystyka systemu APIS. A short analysis of the APIS system. COKB Gdańsk, 1968. CIINTE Zakł. Szkol. 1968
4. A. Targowski, Ocena krajowych systemów automatycznego przetwarzania danych ze szczególnym uwzględnieniem wybranych systemów. The evaluation of national systems of data processing with special referrence to chosen systems. «Maszyny matem.», 1968, № 10
5. J. Wierzbowski, System wyszukiwania informacji IBIS. The IBIS system of information retrieval. IMM Warszawa, 1968. Zakł. Szkol. CIINTE
6. R. Zając, System IGA — automatyczne sporządzanie katalogów. The IGA system — automatic formation of catalogues. Warszawa, 1968. Zakł. Szkolenia
7. J. Bobrowski, Wykorzystanie techniki elektronicznego przetwarzania danych w procesach informacji naukowo — technicznej i ekonomicznej w Polsce. Przegląd zastosowań. The use of electronic

data processing in scientific, technical and economic information in Poland. Review of applications. Warszawa, 1969, CIINTE pp. 40, bibliogr. items 9, Ser. PSP, № 7

8. J. Bobrowski and A. Więckowski, Wprowadzenie emc do procesów informacyjnych w przemyśle elektronicznym. The use of computers for information processes in the electronic industry. CIINTE 1969 APID, № 5, 10—19
9. J. Choroszuszyna and T. Markowski, Indeks permutacyjny typu KWOC do wydawnictwa PPZI APID. The permutationd index of the KWOC type for PPZI APID. CIINTE, 1969, № 4, 9—12
10. J. Wakan, Opis systemu automatycznego wyszukiwania informacji patentowej SAWI — 2P. The description of the automatic system for retrieving the patent information. CIINTE 1970 APID, № 4

В печать 16/VIII-1972 г. Формат бумаги $84\times108^{1}/_{32}$ Тираж 700 экз.
Печ. л. 15 Уч. изд. л. 11,82 Цена 88 коп. Заказ 6138

Производственно-издательский комбинат ВИНИТИ, Люберцы, Октябрьский пр. 403

Errata

page	line	should read
14	8 from top	product
18	13 from bottom	The paper by
53	title	Some Sociological Aspects
89	25 from bottom	distribution of
142	4 from top	what he
162	6 from top	5c1
163	4 from bottom	5c2
221	7 from bottom	entering it
3	5-6 from bottom	In November I962-58, in November I964-58, in I966-6I, in I969-43